CONTENTS

Y.M. Wara & D. Singh
Directorate of Information and Communication Technology, Usmanu Danfodiyo University, Sokoto Nigeria

9-16 An Integration of Clustering And Classification Technique In Software Error Detection.
D.C. Yadav & Pal, S.
Shri Venkatwthwara University, Amroha, India

17-34 A Dynamic Database-Driven Web-Based Automatic NYSC Posting System
V.A. Akpan, O.O. Obe, R.O. Vincent-Akpan & R.O. A. Osakwe
Department of Physics Electronics, Federal University of Technology, Akure, Nigeria

35-40 Enhancement in Security by Reducing Dimensions of Hyperspectral Face Images for Face Recognition
S. Arya, N.Pratap & V. Kumar
Department of Computer Science, Gurukula KangrimVishwavidyalaya, Haridwar Neelkanth Institute of Technology, Meerut

41-52 Multi-agent Based Course Allocator Using GAIA Methodology and JADE Framework.
K.T. Igulu, Z.P. Piah & .P. O. Asagba PhD
Department of Computer Science, Rivers State Polytechnic,Bori, Rivers State, Nigeria.

53-60 An Investigation into the effect of channel estimation and interpolation techniques in long term evaluation system through realistic channel conditions
M. Setia & J. Malhotra
ECE Department, GNDU-RC, India

61-72 A Mobile Architecture For Childhood Vaccine Preventable Illnesses
K. Obahiagbon & B.B. Odigie
Department of Mathematics and Computer science
Benson Idahosa University, Edo State, Nigeri.

73-80 Fish Classification Using Support Vector Machine
S.O. Ogunlana, O. Olabode, S.A. A. Oluwadare & G. B. Iwasokun
Department of Computer Science, Federal University of Technology, Akue, Nigeria

81-90 Performance Evaluation of Routing Protocols in MANETs using Varying Number of Nodes and Different Metrics
O.O. Omitola
Department of Electrical/Electronic and Computer Engineering
Afe Babalola University, Ado Ekiti, Nigeria

91-96 Multi-Agent Based Patient Scheduling Using Ant Colony Optimization
A. A. Obiniyi
Department of Mathematics, Ahmadu Bello University, Zaria, Nigeria

97-106 A New Approach: Automatically Identify Proper Noun from Bengali Sentence for Universal Networking
M. S. Islam & J.K. Das
Department of Computer Science and Engineering
Jahangirnagar University, Savar, Dhaka-1213, Bangladesh

107-114 Configuration, Interfacing and Networking of Wireless IP-Based Camera for Real-Time Security Surveillance Systems Design
V.A. Akpan, R.A.O. Osakwe & S.A. Ekong
Department of Physics Electronics
Federal University of Technology, Akure, Ondo State, Nigeria.

115-120 An Investigation into the Conceptual Controversies between Artificial Intelligence and Computational Intelligence
K.T. Igulu, Z.P. Piah & P.O. Asagba PhD
Department of Computer Science
Rivers State Polytechnic,Bori, Rivers State, Nigeria.

121-126 A Real-Time Voice-Control Model to Access Protected Resources
W.O. Adesanya, O. B. Alaba & G. O. Odulaja
General Studies Department, Federal College of Agriculture, Akure, Nigeria,
Department of Computer Science, Tai Solarin University of Education, Ijagun, Ijebu-Ode. Nigeria
127-132 E-Learning Functional Model: A Technology-Based Teaching Method For Providing Access To Sustainable Quality Education
V.I. Osubor & S.C. Chiemeke
Department of Computer Science, University of Benin, Benin City, Nigeria.

133-144 GOEmbed: A Smart SMS-SQL Database Management System for Low-Cost Microcontrollers
N.E. Osegi & P.Enyindah
Department of Information and Communication Technology
National Open University of Nigeria, Lagos State, Nigeria

145-152 Towards the Development of a Time-Out Multiple C-R CAPTCHA Framework Using Integrated Mathematical Modeling
J.O. Okesola, O.B. Longe & A.P. Obi
School of Computing, University of South Africa, South Africa

153-160 Survey of Opinion of Marginalized College Students About Online University Education
A.O. Itegboje, J.A. Adedokun & O. O. Olusanya
Yaba College of Technology, Yaba Lagos, Nigeria

161-168 Load Balancing Adaption of Some Evolutionary Algorithms In Cloud Computing Environment
A.P. Priyadarshini, R. Sharmistha & K.P. Prasant
School of Computer Engineering, KIIT University, Bhubaneswar, India

169-172 Social Engineering Threats and Applicable Countermeasures
A.O. Adewole & A.E Durosinmi
Department of Computer Science, The Polytechnic, Ibadan, Ibadan, Nigeria.

173-182 Intelligent Allocation of E-Learning Resources Using Tokenization and Crawling Algorithms
Y. Folajimi & E.E. Edet
Department of Computer Science, University of Ibadan, Ibadan, Nigeria

183-194 Harmonic Analysis in a 33kV Distribution Network: A Case Study of Island Business District
L.M. Adesina & O.A. Fakolujo
Department of Inspection and Quality Assurance, Eko Electricity Distribution Plc (EKEDP), Lagos – Nigeria

195-198 Secured Key Management Technique to Detect Blackhole Attack in Mobile Adhoc Network
O.A. Ojesanmi, O.C. Akinola, O. Oluwasusi & S. Salako
Department of Computer Science, Federal University of Agriculture, Abeokuta, Nigeria.
Editorial Board

Editor-in-Chief

Prof. Dele Oluwade
Senior Member (IEEE) & Chair IEEE Nigeria – Computer Chapter.

Editorial Advisory Board

Prof. Gloria Chukwudebe - Senior Member & Chairman IEEE Nigeria Section
Engr. Tunde Salihu – Senior Member & Former Chairman IEEE Nigeria Section
Prof. Adenike Osofisan - University of Ibadan, Nigeria
Prof. Amos David – Universite Nancy2, France
Prof. Clement K. Dzidonu – President Accra Institute of Technology, Ghana
Prof. Adebayo Adeyemi – Vice Chancellor, Bells University, Nigeria
Prof. S.C. Chiemeke – University of Benin, Nigeria
Prof. Akaro Ibrahim Mainoma – DVC (Admin) Nasarawa State University, Nigeria
Dr. Richard Boateng – University of Ghana, Ghana.
Prof. Lynette Kvassny – Pennsylvania State University, USA
Prof. C.K. Ayo – Covenant University, Nigeria
Dr. Williams Obiozor – Bloomsburg University of Pennsylvania, USA
Prof Enoh Tangjong – University of Beau, Cameroon
Prof. Sulayman Sowe, United Nations University Institute of Advanced Studies, Japan
Dr. John Effah, University of Ghana Business School, Ghana

Managing/Production Editor

Prof Longe Olumide
Distinguished Fulbright Alumnus
Member, IEEE Nigeria, Computer Chapter
Foreword

The African Journal of Computing & ICT remains at the nexus of providing a platform for contributions to discourses, developments, growth and implementation of Computing and ICT initiatives by providing an avenue for scholars from the developing countries and other nations across the world to contribute to the solution paradigm through timely dissemination of research findings as well as new insights into how to identify and mitigate possible unintended consequences of ICTs. Published papers presented in this volume provide distinctive perspective on practical issues, opportunities and dimensions to the possibilities that ICTs offer the African Society and humanity at large. Of note are the increasing multi-disciplinary flavours now being demonstrated by authors collaborating to publish papers that reflect the beauty of synergistic academic and purpose-driven research. Obviously, these developments will drive growth and development in ICTs in Africa.

The Volume 8, No. 1, March, 2015 Edition of the African Journal of Computing & ICTs contains journal articles with a variety of perspective on theoretical and practical research conducted by well-grounded scholars within the sphere of computer science, information systems, computer engineering, electronic and communication, information technology and allied fields across the globe. While welcoming you to peruse this volume of the African Journal of Computing and ICTs, we encourage you to submit your manuscript for consideration in future issues of the Journal.

We welcome comments, rejoinders, replication studies and notes from readers.

Very best compliments for the season

Thank you

Longe Olumide Babatope PhD
Managing Editor
Afr J Comp & ICTs
March, 2015
A Guide to Establishing Computer Security Incident Response Team (CSIRT) For National Research and Education Network (NREN)

Y.M. Wara
Directorate of Information and Communication Technology
Usmanu Danfodiyo University
Sokoto Nigeria
E-mail: mywarra@udusok.edu.ng

D. Singh.
Faculty of Information Science and Technology
Universiti Kebangsaan
Malaysia
E-mail: dalbir@ukm.edu.my

ABSTRACT

Academic institutions worldwide embrace communication network and Internet services for their academic activities and collaboration. Communication network enable members of institutions to collaborate with each other and as well access Internet. Therefore, sophisticated cyber-attack emanating from one segment of the network or from the greater Internet can undoubtedly trigger crisis that can affect various stakeholders both within and outside the academic institutions. Consequently, a coordinated concerted effort is required for a more protected network and safe internet. This joint effort is mostly referred to as CERT (Computer Emergency Response Team) or CSIRT (Computer Security Incident Response Team). This paper provides the guides on establishing Computer Security Incident Response Teams (CSIRTs) for securing the National Research and Education Network (NREN) in Nigeria. Focused synthesis method using published articles and archival data has been collected and analysed in this study. Finally, the proposed guide would improve the management of cyber security issues in NREN.

Keyword: CSIRT; Nigeria; Network; Security; Research

I. INTRODUCTION

The advents of communication network and Internet services have turned the world in to a virtual environment or even a global village. Virtual world connects organization’s critical infrastructure to both public and private sectors and as well as academic institutions. A sophisticated cyber-attack can undoubtedly trigger crisis that affect various stakeholders both within and outside the organization. Consequently, numerous communities started to cooperate for a more protected and safe internet. These joint efforts are mostly referred to as CERT (Computer Emergency Response Team) or CSIRT (Computer Security Incident Response Team).

A CERT or CSIRT is an organization of devoted and responsible cyber security experts that plans and responds to cyber security incidents. The organization is responsible for taking, studying, organizing and responding to cyber security incidents and activities [1]. This is very critical since communication networks and information systems have turned out to be crucial element in academics, economics and other social services and development. Therefore, the safety and security of these systems is of high concern to the organization.

To this end, it is imperative for developing National Research and Education networks in Nigeria and the Africa in general to implement superior cyber security mitigation and practices that can efficiently respond to the ever growing information systems attacks through developing plans and supporting resources. One kind of this plans is to have an organize Cyber Security Incident Response Team (CSIRT).

The purpose of this article is to review the steps on establishing Computer Security Incident Response Teams (CSIRTs) for securing the National Research and Education Network (NREN) [2].

This article comprises of four sections, whereby section one explains the concept of NRENs, section two focuses on cyber security incidents, section three reviews on attributes of the CSIRT and section four explains on the application of CSIRT framework in a typical developing NRENs. The contribution of this paper is to guide the development of NRENs in Nigeria and Africa in general to implement a Cyber security incident handling and response team by adopting the guidelines stated herein.
2. NATIONAL RESEARCH AND EDUCATION NETWORK (NREN)

Over the years Research and Education Network (REN) have matured significantly in various countries as a sustainable solution for ICT needs of Higher Education Institutions (HEI). A REN organization generally owns and administers a high performance national backbone network interconnecting all member campuses. National Research and Education Network (NREN) is the interconnection of researched and teaching focused institutions that centred on making use of scarce resources to offer solutions for the purpose of realizing their set organizational goals [3].

Having the need for seamless interconnectivity and interoperability within any network does not warrant compromising security within the Network. There are different levels of security for each user set. For every network on which the national information infrastructure is built must have series of security policies and procedures implemented that will enforce authorized access to the network and any system within it [3]. Networks and its related systems services are considered critical in supporting dispersed research organizations around the globe [3]. In this regard, academic network in African countries can utilize the advancement of these network services to get connected to other NREN around the world for improve academic services.

3. COMMUNICATION NETWORK AND NREN TECHNOLOGIES

It is of high important to perceive that even in the present day, Internet has been feasible to have a perfect combination of enterprise, academics, government and military systems. For sure, the more such diverse organizations continue to connect to the Internet, the more on the increasing opportunity of these intermediate level network systems to have a number of unrestricted clients [3]. Presently, the Internet is properly serving the public but widely vulnerable to mass attacks. This is because Internet is a public network with a wide access to a very large variety of users. It is imperative to know that, as time goes on; many institutional users will support local area networks and will use this network to have access to NREN and as well the Internet. According to [3] there are presently some NRENs that provide research network platform, contents sharing among member institutions and as well provide Internet services without necessarily providing cyber security incident response capability.

Such networks include such as South African National Research Network (SANReN), whereby it was meant to provide Internet services and global research network to south Africa; TENET (Tertiary Education and Research Network of South Africa), is a platform that provides Internet and ICT services to higher educational institutions in south Africa; The Ghanaian Academic and Research (GARNET), is supporting the research and education communities within Ghana promoting academic collaboration; Research and Education Network for Uganda (RENU) is fostering and nurturing teaching, learning and content development in Uganda. West and Central African Research and Education Network (WACREN) is responsible for giving the necessary support for the creation and sustenance of NRENs in the region.

NgREN (Nigeria Research and Education Network) was designed to provide a platform for all higher education institutions in the country to communicate, collaborate and share resources using ICT to promote research and education in the Nigeria. Despite desires and importance of these networks security issues should not be compromised or taken for granted. Different NRENs require different levels of security but there must be a common organized cyber security incident handling procedure and response to mitigate cyber security issues within and the outside NREN.

4. NIGERIA RESEARCH AND EDUCATION NETWORK (NgREN)

Nigeria Research and Education Network (NgREN) like any other NREN was designed to provide a platform for all higher education institutions in the country to communicate, collaborate and share resources using ICT to promote research and education in the Nigeria. Presently, the network has connected all the 27 Federal Universities, National Universities Commission (NUC) and the Committee of Vice-Chancellors' secretariat to a 10gbps core network and a 155mbps last-mile capacity to each University. A 3*STM-1 bandwidth capacity was also provided as a proof of concept.

Each University has a set of Video Conferencing facilities installed. In addition, several efforts are being put in place to connect the remaining Universities, Polytechnics and Colleges of Education and also to promote inter institutional collaboration on the network. NgREN provides connectivity between Nigerian institutions as well as connectivity with other global Research and Education Networks as shown in Figure 1. It is viewed holistically, as a network of educational networks.
5. CSIRT FUNDING AND SUPPORT

This is the first step in establishing a CSIRT within the organization. In many instances, the individuals serving on the project team as well as intermediate managers must be aware and supportive of the project [1]. The management should demonstrate ability and willingness to promote and move the project forward. Besides demonstrating the willingness to support the project, management committee must be ‘ready to release enough funding for the project [1]. It is must for any new forming and existing CSIRT to secure the top level management support of the funding organization.

Any CSIRT that fails to secure initial and continuous management support of the main organization will end of struggling for recognition from the management [4]. This signifies the commitment of stakeholders to ‘buy into’ the decision of forming the CSIRT, that is, to agree to provide support, especially monetary and other resources. Funding is critical and centred to forming CSIRT. The fund should be obtained for start-up, short and long term plans. The funding would be used not only for start-up costs (software, computing equipment, capital furniture expenditures, supplies, Internet domain registration fees, facilities costs, phones, fax machines) but human resources costs (salaries and benefits).
6. COMPUTER SECURITY INCIDENTS RESPONSE TEAMS

6.1. Description of a CSIRT
A CSIRT is a solid authoritative element or capability (i.e., one or more staff parts) that is saddled out the obligation of providing piece of the incident response management for a specific organization [2]. This means that CSIRT is a unit dedicated to guaranteeing that suitable technology and best systems management practices are utilized to counter attacks on networked environment and in addition to restricting harm and guaranteeing coherence of critical services despite effective attacks. Once an incident happens, participants of a CSIRT can aid its constituency in figuring out what happened and what moves need to be made to remedy the circumstances.

This is particularly vital since communication network systems have turned into an important component in financial, educational and social development. Furthermore, the security of these network systems is of growing concern to society [1].

6.2. The evolution of CSIRT
The first CSIRT was made in response to an incident - the 1988 Morris worm. The worm was made to spread across network abusing different vulnerabilities. Consequently, systems such as email and Internet backbone servers were infected. For this reason many Internet sites were removed from the network and this result to total outage of Internet [2]. This problem made US Defence Advanced Research Projects Agency (DARPA) to believe that lack of correspondence between the organizations within the network to evaluate the severity of the worm brought about the failure to avert the situation at an early stage. For this reason DARPA board resolved that an official coordination method was needed to handle such computer security attacks and related analysis and response. In 1988, DARPA chose the SEI as the home of a coordination centred for Internet security incidents [2].

SEI was charged the responsibility of coordinating security emergencies and building awareness across the Internet community. As a result of this, SEI established the CERT coordination Centre (CERT/CC), which metamorphosed over the years into the CERT Division (www.cert.org). Internet had grown from 60,000 host computer systems as at 1988 to 1.2 billion hosts in July 2014 (see figure 2) as measured by the Internet Systems Consortium (see Internet Domain Survey at the Internet Software Consortium). The number of forming CSIRT teams will be increasing to commensurate the growing Internet community worldwide, servicing varieties different sizes domains from organizations, regions and countries (FIRST.org Inc, 2014). Obviously, such a vast constituency would be in effectively served by a single CSIRT.
7. CSIRT FRAMEWORK

At present, a brief description of CSIRTs and their evolution has been discussed section 6.2 and we are now going to have a close look at the main fundamental subjects that must be considered and resolved by every CSIRT. According to the [4] the framework consists of:

- CSIRT Mission
- Constituency
- Organizational placement
- Relationship to other CERTs

As indicated by [4], these four rudimentary components are directly linked to one another as shown in figure 4. If for example, one of the elements is changed or affected the other ones will also change.

7.1. CSIRT mission

This is typically considered as the organization overall goal and objective. A mission statement can be defined as a way in which the organization communicates the business it is in to the outside world [5]. Mission clarifies the organization’s reason of being. The mission of an organization is a pronouncement which distinguishes it among other organizations, issues its character and announces to the outside world why the organization is carrying out its functions [6]. For a CSIRT, the statement of purpose (mission statement) pronounces which services will be given, and the quality of these services. The overall mission’s statement of each CSIRT within it constituency ought to be about giving a decent security to its constituency. The focus of mission statement of a CSIRT should be the team main functions, which is to be a focal point for preventing, receiving and responding to computer security incidents [7].

7.2. CSIRT constituency

A typical meaning constituency rose as a gathering of individuals with differing qualities who are connected by social ties, offer regular viewpoints, and participate in joint activity in topographical areas or settings. The members varied in the emphasis they set on specific components of the definition. The CSIRT constituency can be described as “a particular organization, region or country the CSIRT is design to provide services” [4]. A CSIRT constituency can be boundless or bounded. Most of CSIRTs today are bounded meaning that they only serve the funding organization. Understanding the CSIRT’s constituency will make the team focus for the needs of organization and assets to be protected [2].

7.3. Place in Organization

Diverse structures are proposed for CSIRT by missions, objectives, necessity and services. Teams may be formed in distinctive units such as business, governmental, military, research work, IT, and have diverse structures. Along these lines, CSIRTs are partitioned by shape, size and function to internal, national CSIRTs, coordination centres, investigation centres, vendor’s team and incidents response service. In first step, CSIRT responses to genuine security incidents and decreases attack effects [8].
7.4. Relationship with other Teams

“The Internet is the realm of CSIRTs, and by extension the world. There are many CERT teams around the world that are being served by a CSIRT” [4]. In this respect, at some level these CSIRTs need to inter-operate so as to accomplish their job. This participation and coordination exertion is at the very heart of the CSIRT framework: simply expressing the mission, defining constituency, and deciding the CSIRT’s place within the organization are not sufficient without likewise covering the coordination issue [4].

7.5. CSIRT services

The CSIRT service is one of the derivatives of mission statement of a CSIRT. CSIRT mission is actualized through the services offered by a team to the team’s constituency [4]. CSIRTs around the world may not provide the same set of services. The services to be provided by the team depend on the desires of the organization, the available technical know-how and funding. CSIRT services vary based on organization, purpose and mission of the team. A CSIRT team must provide an incident handling service before it can be considered as a CSIRT. Figure 5 shows the categories of services that a CSIRT could provide to its constituency [2].

i) Reactive services

Reactive services are the core component of CSIRT work. These services are activated by an action or solicitation, for example, a report of a vulnerable host, wide-spreading malicious code, vulnerable software, or something that was identified by an intrusion detection or logging system. Receptive administrations are the centred segment of CSIRT work [2].

ii) Proactive services

These will enable the CSIRT team to immediately prepare to protect the organization assets in expectation of attacks. Security incidents occurrences can greatly reduce by performing these services [2]. Proactive services meant to improve the security of infrastructure and processes within the organization before any incident or even occurs is detected (CERTC/C, 2005).

iii) Security quality management

These services enhance existing and well-implemented services that are free of incident handling and generally performed by different segments of an organization, for example, the IT, audit, or training development. These services are by and largely proactive yet help by reducing the number of occurrence of incidents [2]. These services are meant to improve the general security of an organization (CERT C/C, 2005).

7.6. Quality assurance

Next to CSIRT Services, another derivative of CSIRT mission is Security Quality Assurance. This involves ensuring security quality is maintained for the entire services provided by the CSIRT. This is done by using feedback, risk assessment and business impact analysis.

7.7. Policies and procedures

Another derivative of internal components of CSIRT framework is policy. “It is important to sustain all CSIRT functions and services be sustained by clear policies and procedures” [2]. Policies are considered the guidelines for a CSIRT. These policies includes security policy, incident handling policy, data sensitivity policy, data storage and handling policy, acceptable use policy and as well as internal and external communication, for example handling the media [4]. Procedures for how the policies of the CSIRT can be conducted should also be put in place.

<table>
<thead>
<tr>
<th>Reactive Services</th>
<th>Proactive Services</th>
<th>Security Quality Management Services</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ Alerts and Warnings</td>
<td>+ Announcements</td>
<td>✓ Risk Analysis</td>
</tr>
<tr>
<td>+ Incident Handling</td>
<td>+ Technology Watch</td>
<td>✓ Business Continuity &amp; Disaster Recovery Planning</td>
</tr>
<tr>
<td>– Incident analysis</td>
<td>+ Security Audit or Assessments</td>
<td>✓ Security Consulting</td>
</tr>
<tr>
<td>– Incident response on site</td>
<td>+ Configuration &amp; Maintenance of Security Tools, Applications, &amp; Infrastructures</td>
<td>✓ Awareness Building</td>
</tr>
<tr>
<td>– Incident response support</td>
<td>+ Development of Security Tools</td>
<td>✓ Education/Training</td>
</tr>
<tr>
<td>– Incident response coordination</td>
<td>+ Intrusion Detection Services</td>
<td>✓ Product Evaluation or Certification</td>
</tr>
<tr>
<td>+ Vulnerability Handling</td>
<td>+ Security-Related Information Dissemination</td>
<td></td>
</tr>
<tr>
<td>– Vulnerability analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>– Vulnerability response</td>
<td></td>
<td></td>
</tr>
<tr>
<td>– Vulnerability response coordination</td>
<td></td>
<td></td>
</tr>
<tr>
<td>+ Artifact Handling</td>
<td></td>
<td></td>
</tr>
<tr>
<td>– Artifact analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>– Artifact response</td>
<td></td>
<td></td>
</tr>
<tr>
<td>– Artifact response coordination</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 5: CSIRT Services list by [2]
8. CSIRTs ORGANIZATIONAL MODELS
Organizational structure is the organization’s internal pattern of relationships, authority, and communication. Similarly, organizational structure is defined as “the network of relationships and roles existing throughout the organization” [9]. An organizational model, is a hierarchical structure, defines framework through its system, including lines of power, communications, obligations and asset allocations. Organizational model is also determined by the organization's objectives and serves as the connection in which processes work and business is carried out. Thus, the organizational model decides the number of employees required and their obliged expertise sets [9].

Diverse CSIRT organizations contains missions, objectives, needs, and services. CSIRT teams can be setup in any type of organization. Consequently, CSIRTs are categorized by nature, scope such as internal or national CSIRTs and the kind of services it provides such as incident response, coordination and analysis and product team [8].

Presently, there is no single way to classify CSIRTs. However, research has indicated that CSIRTs can be classified using different features such as scope, nature and functions etc. But, despite these categorization the CSIRTs roles, mission and responsibility remains the same [8]. In the case of NREN, a combined central coordination and a security team is recommended. This is because of the diverse nature of the network. Central coordination is serving as the central coordination for the network and security team will be providing onsite support and as well track attacks within its own local network and report to the central coordination committee.

9. CONCLUSIONS
Due to sophistication of cyber-attacks and lack of cyber security coordination within the developing NRENs, managers have to strategize to contain these attacks in coordinated and organized way by implementing Cyber Security Incident Response Teams (CSIRT). Thus, article has examined the guidelines for establishing Computer Security Incident Response Teams (CSIRTs) for securing the National Research and Education Network (NREN) as its main contribution.

The importance of this paper lies in the practical guidance on what to do at the planning stage of the CSIRT establishment. Even though the paper is aimed at identifying the critical steps needed to establish a CSIRT within an NgREN, all the steps and guidance are relevant for other NRENs. Having a dedicated centralized coordination for cyber security issues as a point of contact for security incidents within an NREN will help in mitigating and preventing major security issues before it becomes disastrous. The weaknesses of not having a CSIRT structure within interconnected network devices are obvious. The inability to contain an incident can lead to repeated incidents. This continual cycle can only lead to disaster.

Future work is proposed to further refine these guidelines, especially regarding establishment of CSIRT for NREN particularly in Nigeria as the case study.

ACKNOWLEDGEMENT
The creation of this paper required support and guidance of a number of people. I therefore, deem it my major duty to show gratitude to my thesis supervisor who provided me constructive criticisms during the development of this paper. Lastly, I am thankful to National Information Technology Development Agency for providing with full financial sponsorship.
REFERENCES


An Integration of Clustering and Classification Technique in Software Error Detection

D.C. Yadav
Research Scholar,
Shri Venkatwhwara University, Amroha, India
E-mail: dc9532105114@gmail.com

S. Pal
Department of MCA,
VBS Purvanchal University Jaunpur (U.P.), India
E-mail: drsaurabhpal@yahoo.co.in

ABSTRACT

The software project development plays important role in software quality. Measuring software quality in a specific manner such as error estimation and check severity of error which is related to the document bug. Data mining create a supportable platform for software project development by which software engineers easily achieved goal of project quality in given time duration and budget. In this paper we integrate both technique classification and clustering for software error detection. Classification technique analyzes the severity of software defect by J48GRAFT, LAD Tree, and BAYESNET also by Clustering technique measure maximum similar data object in data set within same cluster by K-Means.

Keywords - Data mining; Classification: J48graft, Lad Tree and BayesNet; Clustering: K-Means; Weka.

I. INTRODUCTION

Software testing check the performance of software project in which, software tracker plays a technical role to detect the software defect. The software tracker provides feedback information about bug. And give the technical role in software quality improvement. Software tracker provides the severity of bug fix or not fix in the software. Tracker measuring software quality in a specific manner such as error estimation and check the severity of error which is related to the document is known as document bug also decide the severity of bug. Shi and Harjan [1] in 2007 presented data classification of software defect is very commanding mission in data mining. The categories of software defects in data mining used comparable tree function, rule etc. The purpose of arrangement is to analyze the ideas of each adjustable and allocate those variables to corresponding predefined classes.

Jachyra, Pancerz and Gomula [2] introduced about J48graft. J48graft generates a technical graphical way to describe unrecovered problems of document bug in training set. The J48graft generates decision tree and give the training set but it have some drawback which is recover by J48graft. J48graft classify the environment into multi-dimensional space which is not possible by training set it is also reduce the prediction error. In the Fig. 1 J48graft classify the bug of software defect and easily we analyzed bug by J48graft tree.

Buhmann [3] introduced about Lad tree. Lad tree support in the logical analysis of data. It builds a classifier for binary target variable based on learning a logical expression that can distinguish between positive and negative sample in a data set.
LAD tree analyzed software defect in term of binary but not by any negative pattern is positive and similar; binary points covered by some negative pattern but not covered by positive pattern is negative.

Clustering is depending on maximum the similarity between object in same class and minimizing the similarity objects of different classes. Tiwary and Chaudhary introduced about K-means algorithm. It is a centroid based partitions or cluster. K-means algorithm provide k-cluster on the data set.

Let $d_1, d_2, \ldots, d_n$ are data points of software defect data set. Let $C_i$ denotes cluster number for the $i$th data point. K-means minimize scatter by

$$\sum_{k=1}^{K} \sum_{i \in C_k} (d_i - \mu_k)^2$$

When $C_k$ is the mean vector of the $K$th cluster and $N_k$ is the number of data point in $K$th cluster.

2. RELATED WORKS

Shepperd, Schofield and Kitchenham [6] discussed that need of cost estimation for management and software development organizations and give the idea of prediction also give the methods for estimation. Yadav and Pal [7] use the ID3 decision tree to generate the important rules that can help to predict student enrollment into an academic programme called the Master of Computer Application. The generated tree yields that Bachelor of Science students in mathematics and computer applications will enroll and will likely to perform better as compared to Bachelor of Science students without any background in mathematics.

Pal and Pal [8] conducted study on the student performance based by selecting 200 students from BCA course. By means of ID3, c4.5 and Bagging they find that SSG, HSG, Focc, Fqual and FAIn were highly correlated with the student academic performance. Alsmadi and Magel [9] discussed that how data mining provide facility in new software project its quality, cost and complexity also build a channel between data mining and software engineering.

Boehm et al. [10] discussed that some software companies suffer from some accuracy problems depend on his data set after prediction software company provide new idea to specify project cost schedule and determine staff timetable. Ribu [11] discussed that the need of open source code projects analyzed by prediction and get estimating object oriented software project by case model. Nagwani and Verma [12] discussed that the prediction of software defect(bug) and duration similar bug and bug average in all software summery, by data mining also discuss about software bug.

Hassan [13] discussed that the complex data source(audio, video, text etc.) need more of buffer for processing it does not
support general size and length of buffer. Yadav and Pal [14, 15] discussed the use of different classification algorithms using standard quality of software data sets and compared the accuracy level of each method. Li and Reformate [16] discussed that the software configuration management a system includes documents, software code, status accounting, design model defect tracking and also include revision data. Elcan [17] discussed that COCOMO model pruned accurate cost estimation and there are many thing about cost estimation because in project development involve more variable so COCOMO measure in term effort and metrics. Chang and Chu [18] discussed that for discovering pattern of large database and its variables also relation between them by association rule of data mining. Kotsiantis and Kanellopoulos [19] discussed that high severity defect in software project development and also discussed the pattern provide facility in prediction and associative rule reducing number of pass in database. Pannurat, Kerdprasop and Kerdprasop [20] discussed that association rule provide facility the relationship among large dataset as like software project term hug amount, cost record and helpful in process of project development. Fayyad, Piatesky Shapiro, Smith and Utthurusamy [21] discussed that classification creates a relationship or map between data item and predefined classes. Shtern and Vassillios [22] discussed that in clustering analysis the similar object placed in the same cluster also sorting attribute into group so that the variation between clusters is maximized relative to variation within clusters. Runeson and Nyholm [23] discussed that code duplication is a problem which is language independent. It is appear again and again another problem report in software development and duplication arises using neural language with data mining. Vishal and Gurpreet [24] discussed that data mining analyzing information and research of hidden information from the text in software project development. Lovedeep and Arti [25] data mining provide a specific platform for software engineering in which many task run easily with best quality and reduce the cost and high profile problems. Nayak and Qiu [26] discussed that generally time and cost, related problems arises in software project development these problems mention in problem report, data mining provide help in to reduce problems also classify and reduce another software related bugs.

1) The proposed system will analyze severity of software defects predicts. Predicts categorical class level classifiers based on training set and the values in the class level attribute use the model in classifying new data. We integrate both (classification and clustering) techniques. After combine application of most frequent used clustering (k-means) algorithm with classification (J48GRAFT, LAD TREE and BAYESNET) algorithms, the results were compare and the weka data mining tool was used.

3. METHODOLOGY

Our research approach is to use J48graft, Lad Tree, Bayes Net and K-Means: to model the relationships between themenable properties of a software product and its quality. The research methodology is divided into 5 steps to achieve the desired results:

Step 1: In this step, prepare the data and specify the source of data.

Step 2: In this step select the specific data and transform it into different format by weka.

Step 3: In this step, implement data mining algorithms and checking of all the relevant bugs and errors is perform.

Step 4: The decision is taken on the presence of bugs insource code. If Bug is present then proceed further, otherwise it will stop.

Step 5: In this step, we make clusters of particular bug or error with the help of modified K-Means clustering.

Step 6: We classify the relevant bugs using J48graft, Lad tree and Bayes Net algorithm at particular time, after clustering.

Step 7: At the end, the results are display and evaluated.
3.1 Data Preparation

Table 1: Represents Attributes of Document Bug for computation

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source</td>
<td>Name Of A Project Or Department In MASC That Raises The Problem Report.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dependable Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Severity(1)</td>
<td>Problem Report Is Normal</td>
</tr>
<tr>
<td>Severity(0)</td>
<td>Problem Report Is Serious</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Explanatory Variable</th>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bug</td>
<td>(1=Bug Accepted, 0=Bug Not Accepted)</td>
<td>Describe The Bug Or Defect In The Software</td>
</tr>
<tr>
<td>Class</td>
<td>(0=Sw-Bug, 1=Doc-Bug, 2=Ongang Request, 3=Support, 4=Mistaken, 5=Duplicate)</td>
<td>Category Of Bug Class</td>
</tr>
<tr>
<td>Priority</td>
<td>(0=Not, 1=High, 2=Medium, 3=Low)</td>
<td>Describe Schedule Permit Duration</td>
</tr>
<tr>
<td>Time To Fix</td>
<td>0=Within Two Days, 1=Within One Week, 2=Within Two Week, 3=Within Three Week, 4=Within Four Week, 5=Within Five Week</td>
<td>Take Time Duration In Of Problem Report</td>
</tr>
<tr>
<td>State</td>
<td>(0=Closed, 1=Open, 2=Active, 3=Analyzed, 4=Suspended, 5=Resolved, 6=Feedback)</td>
<td>Status Of Problem Report</td>
</tr>
</tbody>
</table>

A software error arises in problem report and all problem reports can be grouped in two categories: severity and none severity. In severity the data set have no error in software in none severity means a software bug arises which is tracked by GANTS which is a bug tracking system in GNU. It is set on MASC intranet to collect and maintain all problem reports from every department of MASC. The document-bug create in software document categories by class field. Now performing for classification of doc-bug using several standard data mining tasks, data preprocessing, clustering, classification, association and tasks are needed to be done. The database is designed in MS-Excel, MS word 2010 database and database management system to store the collect data. The data is formed according to the required format and structures and data is converted to ARFF (attribute relation file format) format to process in weka. An ARFF file is an ASCII text file that describes a list of instances sharing a set of attributes.

3.2 Data Selection and Transformation

In this step only those fields were selected which were required for data mining. A few derived variables were selected. Where some of the information for the variables was extracted from the database. All the predictor and response variables which were derived from database are given in table 1 for reference. The survey uses status of problem report analysis/non analysis and the operationalization of the survey for items is as follows:

6=Feedback, 5=Resolved, 4=Suspended, 3=Analyzed, 2=Active, 1=Open, 0=Closed

The domain values for some of the variables were defined in the table -1 for the present investigation as follow:
3.3 Data Mining Implementation

Weka is an open-source software that implements a large collection of machine learning algorithms and is widely used in data mining applications. From the above data bug.arff file was created. This file was loaded into weka explorer and analyzes severity of software defects predicts. Predicts categorical class level classifiers based on training set and the values in the class level attribute use the model in classifying new data. We integrate both (classification and clustering) techniques. After combine application of most frequent used clustering (k-means) algorithm with classification (J48GRAFT, LAD TREE and BAYESNET) algorithms, the results were compare and the weka data mining tool was used.

The problem in particular is a comparative study of performance of integrated clustering and classification technique i.e. Simple k-means clustering algorithm integrated with different classifier such as J48graft, LAD tree and BAYESNET by using various parameters of document-bug, data set containing 5 attributes, 61 instances and one class attribute.

3.4 Result and Discussion-

To better understand the importance of the input variables and analyzed and performance of document-bug. In our research evaluating the performance of above integrated techniques. The data set needs to be normalized by which removing missing values from data set and if any null field present then there will be removed by adding zero instead of null. After normalized the integrate technique applied in which the k-means technique applied by which divide dataset into number of clusters.

LogScore MDL: -225.7444329214029
LogScore ENTROPY: -166.1367618909699
LogScore AIC: -196.1367618909699

Time taken to build model: 0 seconds

--- Stratified cross-validation ---
--- Summary ---

Correctly Classified Instances 61 100 %
Incorrectly Classified Instances 0 0 %
Kappa statistic 1
Mean absolute error 0.0002
Root mean squared error 0.0012
Relative absolute error 0.0972 %
Root relative squared error 0.3963 %
Total Number of Instances 61

--- Detailed Accuracy By Class ---

<table>
<thead>
<tr>
<th>Class</th>
<th>TP Rate</th>
<th>FP Rate</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
<th>ROC Area</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>one</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>one</td>
</tr>
<tr>
<td>zero</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>zero</td>
</tr>
</tbody>
</table>

Weighted Avg.

--- Confusion Matrix ---

```
e b  <-- classified as
55 0 | a = one
0 61 | b = zero
```

Figure 4: J48graft tree for error detection
Size of the tree : 7

Time taken to build model: 0.02 seconds

--- Stratified cross-validation ---
--- Summary ---

Correctly Classified Instances 61 100 %
Incorrectly Classified Instances 0 0 %
Kappa statistic 1
Mean absolute error 0
Root mean squared error 0
Relative absolute error 0 %
Root relative squared error 0 %
Total Number of Instances 61

--- Detailed Accuracy By Class ---

<table>
<thead>
<tr>
<th>Class</th>
<th>TP Rate</th>
<th>FP Rate</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
<th>ROC Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>one</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>zero</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Weighted Avg. 1 0 1 1 1 1

--- Confusion Matrix ---

a b <-- classified as
55 0 | a = one
0 61 | b = zero

Figure 5: BayesNet for error detection

=== LAD Tree for error detection ===

Figure 6. LAD Tree for error detection
Euclidean distance is a simple distance measure algorithm which calculates the distance of each data values from centroid. Maximum iteration, it is the value that the maximum number of clustering cycle iterates. Number of clusters is what user needs to choose for dividing the data set after clustering the result set needs to be saved in .arff format for applying the classifying algorithm to make integration.

Table 2: K-Mean Error and Time comparison

<table>
<thead>
<tr>
<th>No. Of clusters</th>
<th>No. Of iterations</th>
<th>Sum of squared errors</th>
<th>Time (seconds)</th>
<th>Data sets</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
<td>12</td>
<td>0</td>
<td>19</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>27</td>
<td>0</td>
<td>39</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>45</td>
<td>0</td>
<td>61</td>
</tr>
</tbody>
</table>

From the above Table 3 it may be observed that k-means+J48GRAFT and k-means + BAYESNET both take less time than K-Means + Lad tree. However K-Means + BayesNet has 0.0 MAE, and 0.0 time taken by k-means+J48GRAFT has 0.0 MAE and 0.02 time taken

Table 3: Represents Error Comparison with K-Means between J48graft, Lad tree and Bayes Net

<table>
<thead>
<tr>
<th>Details</th>
<th>K-Means+J48graft</th>
<th>K-Means+Lad Tree</th>
<th>K-Mean+BayesNet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Absolute Error</td>
<td>0</td>
<td>0.0</td>
<td>.0002</td>
</tr>
<tr>
<td>Root Mean Square Error</td>
<td>0</td>
<td>0.0</td>
<td>0.0012</td>
</tr>
<tr>
<td>Time Taken</td>
<td>0.02</td>
<td>0.03</td>
<td>0</td>
</tr>
</tbody>
</table>

According to test and outcome analysis in our research, it was found that in the integrated approach of clustering and classification, the performance of K-Means + BayesNet is better than other algorithms on the basis of time but on the basis of error k-means perform better than other algorithms.

4. CONCLUSION

In this paper, three different classifier are integrated with the simple k-means clustering algorithm and integration techniques were applied on document bug data set. From the observation and analysis it was concluded that the integration of K-Means + BayesNet have 0.0002 MAE and 0.0012 RMSE error and it also takes 0.0 less time to build the model. So on the basis of time and error we found K-Means + BayesNet is better than other. There is large number of classifiers presents and data mining cluster are present. So the future work will be based on other classifier than can be applied on the real data set and also to apply other data mining tools on the data set such that the best techniques can be identified.

REFERENCES

A Dynamic Database-Driven Web-Based Automatic NYSC Posting System

V.A. Akpan, O.O. Obe & R. O. Vincent-Akpan
Department of Physics Electronics
Department of Computer Science
Bursary Department
The Federal University of Technology
P.M.B. 704 Akure, Ondo State, Nigeria.

R.O. A. Osakwe
Department of Physics
The Federal University of Petroleum Resources
P.M.B. 1221 Effurun, Delta State, Nigeria.

† Corresponding Author, Email: avincent@eng.auth.gr

ABSTRACT

This paper presents comprehensive techniques for the development, implementation and deployment of an automatic dynamic database driven web-based system for posting national youth service corps (NYSC) members based on three major criteria, namely: 1) prospective corp members above the age of 30 years are exempted from the NYSC service programme; 2) prospective corp members who are below the age of 30 years and married are posted to the state where their spouse is located; and 3) prospective corp members below the age of 30 years and single are posted to a state and to a region different from their state of origin and state of residence. The design of the database driven web-based NYSC posting system has been accomplished with the aid of five major software tools, namely: 1) Apache web server for hosting the application, 2) Adobe Dreamweaver for writing HTML for the web-based application, 3) PHP for writing scripts required for the application, 4) MySQL as the database server for writing SQL, 5) Google Chrome browser for reading and displaying information developed using the first four tools. Significant instructions on the installation, setting-up and configuring these five software tools have also been presented which enables these tools to work interactively. The results obtained from the implementation of the proposed dynamic database driven web-based NYSC posting system show the validity and efficiency of the NYSC posting system. Finally, it has been recommended that the proposed NYSC posting system can be deployed to NYSC headquarters for the posting of NYSC corp members to replace and speed up the manual method of posting NYSC corp members.

Keywords – Dynamic database, hypertext mark-up language (html), Java® Script®, MySQL, NYSC Posting, PHP, WAMP server, web development.

I. INTRODUCTION

The time lag, paper work and rigours involved in manual method of posting of Corps members to states of the country makes it more desirable but imperative that computerized approach be used to the facilitate in National Youth Service Corps (NYSC) members postings. The manual method suffers a number of setbacks which includes large volume of requests for redeployments at the National Youth Service Corps (NYSC) headquarters from corps members who have peculiar cases warranting redeployment. The manual process is time consuming and prone to errors. It has led to corps members posting procedure to remain cumbersome over a long period of time and the result of this is that corps members getting their postings at the last minutes.

Corps members have just few hours to a day before getting to their various states of postings. Therefore, the problem that arises is to find a method of processing intending corps members’ personal information while making use of the NYSC scheme criteria for posting that would be sufficiently accurate in posting intending NYSC and within a reasonably time and reducing cost in the long run. The software application in this research is intended to bring relieve by providing for timely, accurate and cost-effective process of posting NYSC Corps members using the processing power of the computer software as well as information and communication technologies (ICT).
The NYSC Scheme was created in a bid to reconstruct, reconcile and rebuild the country after the Nigerian Civil war. NYSC was established in 22nd May 1973 with a view to proper encourage and develop of common ties among the Nigerian youths and the promotion of national unity. At present the National Youth Service Corps has some criteria to be considered before a Corps member is posted to the state of national service. Some of the criteria are: 1). State of origin; 2). State of residence; 3). States in Nigeria that has been visited or lived before? 4). Marital Status; 5). If married, state residence of spouse; 6). Age at the next birthday; 7). Nigerian languages that corp member can speak; 8). State of higher institution; 9). Any health or medical challenges? 10). Can treatment be obtained from any hospital or strictly family institution; 11). Any health or medical challenges? 10). Can treatment be obtained from any hospital or strictly family institution; 11). Are parents aged? and 12). Position in the family? Note that the only child should be posted closer home.

This study attempts to develop a database driven website that can post corps members to states in Nigeria using the NYSC Scheme criteria to replace the manual method of posting of corps members. The database driven website proposed in this study is such that posting of corps members will be done automatically by database driven website thereby reducing time, cost and more accurate posting will be achieved. The requests for redeployments will be reduced. The main focus of this study is on the design, development and implementation of a dynamic database driven web-based system for the posting of prospective NYSC corp members. This proposed automatic NYSC posting system will be accomplished according to the following five sub-objectives:

1. Formulation of the dynamic database driven web-based system problem for the posting of prospective NYSC corp members.
2. Critically specify, install, set-up and configure the software required for the design, development and implementation of a dynamic database driven web-based system.
3. The design and development of a dynamic relational database management system (RDBMS) using MySQL programming via MySQL server.
4. Coding the different procedures for posting of the prospective NYSC corp members as PHP script files.
5. Implementation and validation of the proposed dynamic database driven web-based system for posting NYSC corp members according to the following three main criteria: (i) Prospective corp members above the age of 30 years are exempted from the NYSC service programme; (ii) Prospective corp members who are below the age of 30 years and married are posted to the state where their spouse is located; and (iii) Prospective corp members below the age of 30 years and single are posted to a state and to a region different from their state of origin and state of residence.

In this study, a database driven website has been proposed to solve the problem of posting of corps members considering the three main criteria above for determining the state of posting. A database driven website also called dynamic web pages or websites is a website that grabs information from a database.

The webpage is connected to the database through programming and inserts the information into the web page each time it is loaded. The proposed dynamic database driven web-based system will be designed for posting prospective corp members to different states in different regions at the national level on the basis of the following three criteria:

1). Prospective corp members above the age of 30 years are exempted from the NYSC service programme;
2). Prospective corp members who are below the age of 30 years and married are posted to the state where their spouse is located; and
3). Prospective corp members below the age of 30 years and single are posted to a state and to a region different from their state of origin and state of residence.

The proposed design methodology as well as the NYSC posting system can be adapted for posting corp members within a state to respective private and public sectors, industries and parastatals based on the corp members course of study and area of specialization.

The paper is organized as follows. Section 2 presents the overview of the proposed design workflow. Installation, setting-up and configuration of the software used in this study are presented in Section 3. The development, implementation and deployment of the proposed dynamic database-driven web-based automatic NYSC posting system is detailed in Section 4. Section 5 concludes the paper with some discussions on shortcomings and future directions.

2. OVERVIEW OF THE PROPOSED DESIGN WORKFLOW

Databases are a fantastic way to store most types of relational data and information. A database allows one to store rational information in a logical format making it easy to manage and retrieve. It is possible to take almost any existing database and integrate it into website [1]. A database driven website is a website that uses a database to gather, display, or manipulate information. A database driven website is a website that has most of its web pages content in a database [2]. According to Yank (3), (4), database driven website programming can also be called or characterized as server side programming. The reason it is so called because the action or magic that allows the web pages to connect to the database is actually taking place in the server. The web pages that are created when a database website design solution is used are actually called dynamic web pages.

According to DB Net Solution [1], people have encountered databases such as Microsoft Access or software based on databases. Database allows storage of relational information in a logical format, thus making it easy to manage and retrieve such information. It is also possible to take almost any existing database and integrate it into a website. Quackit [2] stated that some websites have a combination of static and dynamic content. The reason for this is that smaller website will be static and vice versa. There is little need to configure a database just to store a handful of web pages; it is much easier and cheaper to keep such pages as files on the server.

According to DB Net Solution [1], people have encountered databases such as Microsoft Access or software based on databases. Database allows storage of relational information in a logical format, thus making it easy to manage and retrieve such information. It is also possible to take almost any existing database and integrate it into a website. Quackit [2] stated that some websites have a combination of static and dynamic content. The reason for this is that smaller website will be static and vice versa. There is little need to configure a database just to store a handful of web pages; it is much easier and cheaper to keep such pages as files on the server.
Today, websites are often developed using dynamic pages driven by databases. The web pages that are created when a database websites design solution is used are actually called dynamic web pages. A website with dynamic content usually has a CMS (content management system) to assist the content providers in updating the website. According to Quackit, (2012), a CMS is usually provided in the form of an administration area where content providers need to log in before they can add contents. The information contained in the databases is usually kept up to date by using a website content management system. The content management system can be designed and tailored to specific requirement. Once logged in, they can create, update and delete articles. They may be able to upload files such as word documents, PDF files, etc. They may be able to upload images too. All this content can be stored in the database. Some may be stored on the file system too. An example, are documents and images that can be stored in the database, they are sometimes stored on the file system as performance and database size are key reasons.

2.1 Reasons for Using Database Driven Web Pages

According to Rogers [5], the following are the reasons for using database driven web pages: 1). Storing and distribution of large amounts of information (e.g. Library/information centers, news articles, etc.); 2). The need to add, update, maintain and display historical data (e.g transactional data entry systems, surveys, polls, customer information); 3). To manage membership-based or organization. Storing information about members, such as their bio and profiles, creating membership directories and implement the appropriate security requirements in members only content areas; 4). For the maintenance of large volume of static web pages; 5). For centralization of data; 6). To share the same content across several different pages. Storing that information a database ensures that only the database table is updated; and 7). For more effective website maintenance and content management.

2.2 Advantages of a Database Driven Website

In the following we highlight some advantages of the proposed database driven website:

1). Time: Managing a website can become time consuming and costly, especially if a large number of services is to be offered on the website. It is simpler if each of this service occupies a single record in a database. Once a service is changed the database can simply be uploaded unto the website, and the web pages are updated in one simple action.

2). Skills: No HTML or website design knowledge is required. Using database completing simple forms website can easily be updated. Large complex websites can be maintained by user with very little computer experience.

3). Cost: Using database driven website is cost effective. Content management tools give freedom to create dynamic web page content, rather than being dependent upon a third party, a web designer.

4). Interactivity: Database driven website present visitors with information which relates specifically to individual needs. It also provides interactive elements to ensure that users have reasons to return to the website on an ongoing basis.

5). Ability to evolve: Website has the ability to evolve without incurring extra costs. Website can be as large as possible. Just by adding more records to the database extra web pages will automatically be created.

6). Databases add a new dynamic angle to the website giving additional leverage and accessibility to content providing a richer and more interactive user experience and at the same time reducing the support and maintenance overhead associated with maintaining static websites.

7). Updating: Incorporating a database into websites can reduce the hassle and complexity of maintaining and updating the website. It is highly searchable and organized making content far more accessible to the end user. Traditional software based systems can be migrated.

2.3 An Overview of a Database Driven Web-based Design

Web applications are by nature distributed applications. In particular, this means that one part of the application is executed on the web server while another runs on the client computer within a web browser window as illustrated in Fig. 1. When the user types in a web address into the URL (i.e. uniform resource locator) and hits the enter key (or clicks on a link or chooses a bookmark), the web browser sends this address into the so-called HTTP (HyperText Transfer Protocol) request to the web server. If the web server receives a request for a static HTML file (typically indicated by a file name ending in .html or .htm) it merely locates the corresponding file on its hard-drive and sends the requested information back to the computer that had requested it. If the web server receives a request for dynamic files (often indicated by .php, .asp, .cfm, .jsp etc.), it does a little bit manipulation before sending the response back to the web browser.

Fig. 1: A web browser requests a document from the web server. The web server parses and executes a PHP script which dynamically assembles the HTML code that is then sent back to the browser which renders the HTML code into a webpage.
2.3.1 The Main Task of the Web Browser

The web browser is responsible for displaying the layout described by the HTML file it has received (this process is called “page rendering”). Once the page is rendered on the screen the user can interact with it; it involves clicking on the links, fill in forms etc. Since HTML alone provides only a crude level of interactivity (namely links, form elements, buttons and image maps), JavaScript was developed. JavaScript is a programming language which gives the web designer a number of options for creating sophisticated user interfaces (e.g., input validation, pop-up windows, cascading menus using cascaded style sheet, CSS). JavaScript code is often contained within an HTML file enclosed by the `<script>` tag. JavaScript is always executed by the web browser.

The web server is fully ignorant about JavaScript (although, there is something like server-side JavaScript but this subject is outside the scope of this study and we refer interested readers to [6]–[8]. However, the web server just serves up any HTML, JavaScript or pictures that have been requested without analyzing or executing any code. For clarity and to avoid confusion; beyond its name, JavaScript has very little in common with Java, which is a programming language that is often used for programs (such as java servlets, java server pages, e.t.c.) running on the web server as well as for highly-interactive (but sometimes slow and crash-prone) java applets running on the client computer. Therefore, when creating a mental map of web technologies, it is better to place JavaScript and Java in opposite corners.

2.3.2 The Main Task of the Web Server

So far, we have only discussed how a web server handles requests for static files (HTML pages, images, etc.). However, if we want to construct our web pages from data contained within a database, the web server performs additional tasks. If the web server receives a request for a dynamic file (indicated by a file ending .php, .asp, .cfm, .jsp, .etc.) running on the web server as well as for highly-interactive (but sometimes slow and crash-prone) java applets running on the client computer. Therefore, when creating a mental map of web technologies, it is better to place JavaScript and Java in opposite corners.

Typically PHP commands output certain HTML code depending on user input or data read from a database etc. This resulting HTML code is then sent back to the web browser for rendering.

2.4 Setting up a Database for the Database Driven Web-Based Applications

According to Jochen [9], databases store data persistently. A database is a piece of software which often runs on the same physical computer as the web server. Data can be collected via the web, web applications are by nature distributed applications. In particular, one part of the application is executed on the web server while another runs on the client computer within a web browser window. Before building a dynamic web site, the tools needed are:

1). PHP: This is a server-side scripting language. It is a plug-in web server that enables it to do more than just exact copies of the files that the browsers ask for. With PHP, web server will be able to run little programs call PHP scripts. It can do little tasks like retrieving up to the minute information from a database and use it to generate a web page on the fly before sending it to the browser that requested it. For PHP scripts to retrieve information from a database, a database must exist. It processes the page request and fetches the data from the MySQL database, then spits it out dynamically as the nicely-formatted HTML page that the browser expects [3], [4].

2). MySQL: This is a relational database management system or RDBMS. It is a software program that is able to organize and manage many pieces of information efficiently while keeping track of how all of those pieces of information are related to each other. It also makes that information really easy to access with server-side scripting languages like PHP.

3). WampServer: There WAMP stands for Windows, Apache, MYSQL and PHP is an all-in-one program that included built-in copies of recent versions of the Apache web server, PHP and MySQL.

4). Opera Browser: This is used to interpret the codes written in HTML and other scripting languages, the browser is a used to read and give visual interpretation to the html codes.

The whole idea of a database driven website is to allow the content of the site to reside in a database, and for the content to be dynamically pulled from the database to create web pages for viewing purposes with a regular web browser. The whole idea of what happens when a page on database-driven website is visited is illustrated in Fig. 2. Thus, at one end of the system it is possible to have a visitor to the website that uses a web browser to load, say for example http://www.abc.com/helloworld.php, and expect to view a standard HTML web page. At the other end, it is possible to have the content of the website, which sits in one or more tables in MySQL database that only understands how to respond to SQL queries (commands).

As shown in Fig. 2, the PHP scripting language is the go-between that speaks both languages. It processes the page request and fetches the data from MySQL database, then outputs dynamic information as nicely-formatted HTML page that the browser expects. With PHP, presentation aspects
of the site can be written (the fancy graphics and page layouts) as templates in the HTML. Where the content belongs in those templates, PHP code is used to connect to the MySQL database and uses SQL queries. According to Yank [3], [4], the following happens when a database-driven website is visited:

1) The visitor’s web browser requests the web page using a standard URL.

2) The web server software (Apache, IIS, or whatever) recognizes that the requested file is a PHP script, and so the server interprets the file using its PHP plug-in, before responding to the page request.

3) Certain PHP commands connect to the MySQL

4) The MySQL database responds by sending the requested content to the PHP script.

5) The PHP script stores the content into one or more PHP variables, and then uses the now familiar echo function to output the content as part of the web page.

6) The PHP plug-in finishes up by handing a copy of the HTML it has created to the Web server.

7) The Web server sends the HTML to the Web browser as it would a plain HTML file, except that instead of coming directly from an HTML file, the page is the output provided by the PHP plug-in.

3. INSTALLING, SETTING-UP AND CONFIGURING THE SOFTWARE FOR THE PROPOSED DESIGN WORKFLOW

In the current workflow, five major software are employed for the development of the database driven web-based NYSC posting system, namely: 1) Apache web server for hosting the application, 2) Adobe Dreamweaver for writing HTML for the web-based application, 3) PHP for writing scripts required for the application, 4) MySQL as the database server for writing SQL, 5) Google Chrome browser for reading and displaying information developed using the first four tools. These software were installed on a Dual Core Pentium CPU computer having 320GB of hard disk with 2.00GB RAM running Microsoft® Windows® 7 operating system.

3.1 Installing, Setting-Up and Configuring the Software Required for the Development of the Database Driven Web-Based NYSC Posting System

3.1.1 Setting-Up and Configuring the PHP

1) Open the php.ini file with a favorite text editor. If there is no particular preference, just double-click the file to open it in Notepad. It’s a large file with a lot of confusing options, but look for the line that begins with doc_root (Notepad’s Edit > Find… feature will help). Out of the box, this line looks like:

\[
\text{doc_root} =
\]

At the end of this line, add the path to the web server’s document root directory. For the Apache server, this is the htdocs folder in the main Apache web server directory. If Apache is installed in the default location, the path should be “C:\Program Files\Apache Software Foundation\Apache2.2\htdocs”. If Apache is installed elsewhere, find the htdocs folder and type its path:

\[
\text{doc_root} = \text{“C:\wamp\bin\apache\Apache2.2.1\htdocs”}
\]

2) Just a little further down in the file, look for the line that begins with extension_dir, and set it so that it points to the ext subfolder in the PHP folder:

\[
\text{extension_dir} = \text{“C:\PHPext”}
\]

3) Scrolling down further in the file, and one will find a bunch of lines beginning with :extension=. These are optional extensions to PHP, disabled by default. We want to enable the MySQL extension so that PHP can communicate with MySQL. To do this, we remove the semicolon from the start of the php_mysqli.dll line becomes:

\[
\text{extension=php_mysqli.dll}
\]

Note: Please note that we are referring to “php_mysqli” and not “php_mysql“. Just above the line for php_mysqli.dll there is a line for php_mysql.dll. The i in php_mysql stands for improved. Here, we want to enable this new improved MySQL extension. The one without the i is obsolete, and so some of its features are incompatible with current versions of MySQL.

4) Scrolling down even further down in the file, and looking for a line that starts with :session.save_path. Once again, we remove the semicolon to enable this line, and set it to the current Windows Temp folder to have the following:

\[
\text{session.save_path} = \text{“C:\Windows\Temp”}
\]

Save the changes made and we close the text editor.

5) The above tasks takes care of setting-up and configuring of the PHP.

3.1.2 Setting-Up and Configuring the Apache Web Server to PHP as a Plug-in

1) Run Notepad as Administrator. This is necessary because the Apache configuration file, by default, can only be edited by an administrator. To do this, find the Notepad icon in the Start Menu (under All Programs > Accessories) and right-click on it. Click the Run as administrator menu item.

2) Choose File > Open in Notepad. Browse to the conf subfolder in the Apache installation folder (C:\wamp\bin\apache\Apache2.2.1\conf), and select the httpd.conf file located there. In order to make this file visible for selection, it may be necessary to select All Files from the file type drop-down menu at the bottom of the Open window.

3) Look for the existing line in this file that begins with DirectoryIndex. Shown here:

\[
<\text{IfModule dir_module}>
\text{DirectoryIndex index.html}
<\text{IfModule}>
\]

This line tells Apache which filenames to use when it looks for the default page for a given directory. Add index.php to the end of this line:

\[
<\text{IfModule dir_module}>
\text{DirectoryIndex index.html index.php}
<\text{IfModule}>
\]

4) All of the remaining options in this long and intimidating configuration file should have been set up correctly by the Apache install program. Next we need to add the following lines to the very end of the file:

\[
\text{LoadModule php5_module “C:\PHP\php5apache2_2.dll”}
\]<httpd-php.php
\text{PHPIniDir “C:\PHP”}
\]
7). Security Options: Uncheck the Modify Security Settings option. It’s best to learn how to set the root password mentioned at this juncture without the assistance of the wizard, so we will show how to do this in the section called “Post-Installation Set-up Tasks”. Once the wizard has been completed, the system should now be fully equipped with for running MySQL server! To verify that the MySQL server is running properly, we use the Ctrl+Alt+Del tabs and choose the option to open the Task Manager. Then click the Show Processes from all users button unless it has already been selected. If all is well, the server program (mysqld.exe) should be listed on the Processes tab. It will also start up automatically whenever the system is restarted.

To add the MySQL command prompt programs that come with the Wamp Server to the Windows system path, we perform the following tasks:

1). Open the Windows Control Panel. Locate and double-click the System icon.
2). In our Windows 7, we click the “Advanced System” settings link in the sidebar.
3). Click the Environment Variables… button.
4). In the list labeled User variables for user, look for a variable named PATH.
   (i) If it exists, select it and click the Edit… button.
   (ii) If there’s no variable, click the New… button and fill in the Variable name by typing PATH.
5). Add the path to Wamp Server’s MySQL bin directory as the Variable value:
   (i) If the Variable value is empty, just type in the path. In our case, the path is:
       “C:/wamp/bin/mysql/mysql5.1.36/bin”
   (ii) If there is already text in the Variable value field, add a semicolon (;) to the end of the value, we type the path name thereafter. If the Variable value field does not exist, it must be created.
6). Click the OK button in each of the open windows to apply and effect the changes.

3.2 Implementation Steps: The Research Methodology for Design of the Database Driven Web-Based NYSC Posting System

Step 1: Statement of the Problem

In order to reduce data and information redundancy, the very first step involves creating a layout of the overall design for the NYSC posting problem based on some criteria such as: (i) State of origin; (ii) State of residence; (iii) States in Nigeria that has been visited or lived before?; (iv) Marital Status; (v) If married, state residence of spouse; (vi) Age at the next birthday; (vii) Highest level of education obtained; (viii) State of higher institution; (ix) Any health or medical challenges? If yes, can treatment be obtained from any hospital or strictly family treatment? (x) Are parents aged? and (xii) Position in the family? In such a case, the only child should be posted closer home.

In the current workflow, three criteria are implemented for posting an NYSC corp member. The three criteria are as follows:

1). A prospective NYSC corp member that is above 30 years of age should be exempted.

3.1.3 Setting-Up and Configuring MySQL Database Server

1). Server Type: Assume that one is setting up MySQL for development purposes on a desktop computer, choose Developer Machine.
2). Database Usage: Unless we know that we will need support for transactions (as such support is usually superfluous for most PHP applications), then we choose Non-transactional Database Only.
3). Connection Limit: Select Decision Support (DSS)/OLAP to optimize MySQL for a relatively modest number of connections.
4). Networking Options: Uncheck the Enable Strict Mode option to ensure MySQL’s compatibility with older PHP code that might be needed for use in that case.
5). Default Character Set: Select Best Support For Multilingualism to tell MySQL to assume that we want to use UTF-8 encoded text, which supports the full range of characters that are in use on the Web today.
6). Windows Options: Allow MySQL to be installed as a Windows Service that's launched automatically; also select Include Bin Directory in Windows PATH to make it easier to run MySQL’s administration tools from the command prompt.

Fig. 3: The PHP version number indicates Apache is configured to support PHP.

We make sure the LoadModule and PHPIniDir lines point to the PHP installation directory, and note the use of forward slashes (/) instead of backslashes (\) in the paths as well as in the path names.

5). Save all changes and close Notepad.
6). Restart Apache using the Apache Service Monitor system tray icon. If all is well, Apache will start up again without complaint.
7). Double-click the Apache Service Monitor icon to open the Apache Service Monitor window. If PHP is installed correctly, the status bar of this window should indicate the version of PHP that have been installed, as shown in Fig. 3.
8). Click OK to close the Apache Service Monitor window.
2). A prospective NYSC corp member that is married and below 30 year of age should be posted to the state where their spouse is located.

3). A prospective NYSC corp member that is single and below 30 year of age should be posted to a state different from: (i) his/her state of origin, (ii) his/her state of residence, and (iii) where he/she studied.

Step 2: Creation of the Database
The second step is to design the database. In the current workflow, MySQL software is used to create a relational and dynamic database system named nysc_posts while the main database file is nysc_posts.sql. Here, the database is created from the MySQL console using the following syntax:

```
mysql> CREATE DATABASE nysc_posts;
```

All subsequent development within the database is done from the MySQL graphical user interface (GUI). Using the MySQL GUI different character types, lengths, as well as other properties are specified for the different fields.

Step 3: Writing the PHP Script Files
The third step is to systematically write codes to implement the NYSC posting system. The term “systematically” used here implies that the HTML codes are written while PHP functions are inserted accordingly which will link the HTML code pages to the database in a cascaded format using the cascaded style sheet (CSS) tool properties. In the current workflow, Adobe Dreamweaver™ is used.

Step 4: Deploying and Viewing the Design in a Google Chrome Web Browser
The fourth step is to view the codes in a browser. When the URL (uniform resource locator) is set to the domain name, the webpage comes up from where the user can enter in the required information. The browser is actually used to interpret the codes written in HTML and other scripting languages. In the current workflow, Google Chrome Version 23.0.1271.64m 2012 has been used as the browser since it supports JavaScript and HTML version 5.

Step 5: Viewing the Database Information and Status
The last but optional step for verification of the database information may be performed. Here, the database is re-visited in order to view the information that has been collected from the user. At this point as well as at any other point, the information collected can be edited, modified or deleted.

4. DEVELOPMENT, IMPLEMENTATION AND DEPLOYMENT OF THE PROPOSED DATABASE-DRIVEN WEB-BASED AUTOMATIC NYSC POSTING SYSTEM

4.1 Statement of the Problem
In order to reduce data and information redundancy, the very first step involves creating a layout of the overall design for the NYSC posting problem based on some criteria such as: 1). State of origin; 2). State of residence; 3). States in Nigeria that has been visited or lived before?; 4). Marital Status; 5). If married, state residence of spouse; 6). Age at the next birthday; 7). Spoken Nigerian languages; 8). State of higher institution; 9). Any health or medical challenges? If yes, can treatment be obtained from any hospital or strictly family treatment?; 10). Are parents aged?; and 11). Position in the family? Again, in such a case, the only child should be posted closer home to his/her parents.

Thus, in the current workflow, three criteria are implemented for posting an NYSC corp member. The three criteria are as follows:

1). A prospective NYSC corp member that is above 30 years of age should be exempted;

2). A prospective NYSC corp member that is married and below 30 year of age should be posted to the state where their spouse is located; and

3). A prospective NYSC corp member that is single and below 30 year of age should be posted to a state different from: (i) his/her state of origin, (ii) his/her state of residence, and (iii) where he/she studied.

Fig. 4: The form for collecting school details.

Fig. 5: Two-field form for administrator’s verification log-in.
4.2 Implementation of the Database Driven Web-Based NYSC Posting System

4.2.1 The Client (User) and Server (Administrator) Login Information

The design and implementation of the proposed database driven web-based NYSC posting system are in two parts similar to the implementation illustrated in Fig. 1, namely: the client (user) side and the server (administrator) side. The user can log-in from any system logging on to the URL of the proposed system (http://localhost/nysc), provided the host can be accessed by the network either via the Internet or a local area network (LAN).

The client (user) side takes the user to the “Process Posting” page shown in Fig. 4. On this page, the user is required to select their appropriate state, school (institution), school authentication number (given in Table A.1 of Appendix A), and finally specify the number of students from their respective school that are to be posted for the NYSC service year (for example 10 or 500 as the case may be). For security and validation of the school authentication, additional verification dialog box is provided. After successfully supplying this information, the user clicks the “Submit” buttons. Clicking the “Submit” buttons automatically takes the user to the registration page shown in Fig. 9 which is discussed in the next section.

On the server (administrator) side, the administrator must log-in to access and process secured information relating to the actual posting of the prospective NYSC corp members. The administrator login-in page is shown in Fig. 5. As it can be seen in Fig. 5, the administrator’s username and password are required. In this workflow, the administrator’s username and password are “admin” and “12345” respectively. After supplying the correct username and password, the administrator clicks the “Login” buttons. Clicking the “Login” buttons automatically takes the administrator to another page which displays the form used for posting NYSC corp members shown in Fig. 6.
Fig. 9: Forms allocated to the user: (a) the accessible form for entering student’s information, and (b) inaccessible forms pending the completion of the first form “Form 1” in (a) and clicking the “Submit” button.

As can be seen on the form in Fig. 6, the administrator must select the state and the school (institution). Once these two items have selected, the “POST” button pops up as shown in Fig. 17 which is discussed in the next section. Note that form used for posting NYSC corp members shown in Fig. 6 also gives the administrator information and instructions on how the posting of NYSC corp members can be done in a simplified manner together with the conditions used for posting prospective NYSC corp members.

4.2.2 Registering NYSC Qualified students
4.2.2.1 Implementing and Processing the Client (User) Information

As discussed in the previous section 4.2.1, the design and implementation of the proposed database driven web-based NYSC posting system are in two parts, namely: the client (user) side and the server (administrator) side. Here, the client (user) log on to the URL of the proposed system (http://localhost/nysc) from a computer denoted as “System A” via an Ethernet-based local area network (LAN) and connects to the server (administrator) host computer denoted as “System B”. This action opens the “Process Posting” page shown previously in Fig. 4.

In order to demonstrate the proposed implementation strategy, Benue State has been selected as the state from where students will be posted with state identification number (state_id) as 9; and the Benue State University has been selected as the school (institution) with the assigned Authentication code as 12362.
Thus, under the “Select State” drop down menu on the “Process Posting” page, Benue State is selected while Benue State University is selected from the “Select School” drop down menu. The school’s authentication code of 12362 (see Table A.1 of Appendix A for school authentication codes) is entered under “Authentication” while 10 has been entered as the number of qualified prospective NYSC corp members to be posted from the Benue State University. The completed form on the Process Posting is shown in Fig. 7.

Note that the Authentication codes are generated by the administrator for each school and are sent secretly to individual school representatives for use during login. For this reason, the authentication code for Benue State University is not visible and it is only valid for the school it has been assigned. As discussed in the previous section, it is necessary to re-validate the authentication code using the second dialog box shown in Fig. 8.
After successfully supplying this information, the “Submit” button is clicked. Once the “Submit” is clicked, the proposed database driven web-based NYSC posting system checks and verify that the information provided are complete and correct. Otherwise, it clears all information entered and takes the client (user) back to the page shown in Fig. 4 so as to refill the required information correctly. On the other hand, if the information entered on the forms in Fig. 7 and Fig. 8 are correct, then clicking the “Submit” buttons automatically takes the user to the registration page shown in Fig. 9. Recall that the school representative entered 10 as the number of students to be posted for NYSC from the Benue State University in Fig. 7. The proposed database driven web-based NYSC posting system automatically created 10 forms for these 10 allocations.

The proposed system sets the first accessible form for the first student on the left as can be seen in Fig. 9(a) while the forms for the remaining 9 students are inaccessible but reserved in the second column as shown in Fig. 9(b). Here, the school representative for Benue State University supplies all the information for the first student as originally supplied by the student to the University. The complete information for the first student is shown in Fig. 10.

Once the submit button is clicked, the student’s information is automatically save and the processed information on the first form for the first student is stored on a newly generated form with the appropriate serial number at the bottom as shown in Fig. 11. As one can observe in Fig. 11, the date of birth for the first student was not entered and the form specified it as “undefined” but accepted.

This kind of mistake generates error during the posting process and should be avoided. Please observe that as the current form is filled and submitted, the next form (in this case Form 2) becomes accessible and so on. Also note that the proposed database driven web-based NYSC posting system automatically calculates and displays the age of a prospective NYSC corp member once the data of birth is supplied. The completed and saved forms for the first two students with correct information are shown in Fig. 12.

The proposed database driven web-based NYSC posting system has also been design so as to inform the user about the success of a particular registration by successively printing on the inaccessible form the name of a registered student once the “Submit” button has been clicked (see for example “Akpan Vincent has been Registered” on the last inaccessible form in Fig. 11, “Bassey Victory has been Registered” in Fig. 12, etc).
The typical layout of the completed and saved students’ registration forms including the third student together with all the registered students, the fourth accessible form and the remaining five inaccessible forms are shown in Fig. 13.

![Completed registration forms](image)

Fig. 15: The form for posting NYSC corp members.

![Administrator dialog box](image)

Fig. 16: Administrator dialog box for selecting a state and school for posting NYSC corp members.

4.3 Posting NYSC Students

4.3.1 Implementing and Processing the Server (Administrator) Information

At the administrative end, although posting is done automatically by the proposed database driven web-based NYSC posting system; only the administrator still has the right to perform the posting of students. The administrator must lunch [http://localhost/nysc/login.php](http://localhost/nysc/login.php) to arrive at the login page shown in Fig. 14. Similar to the discussion in Section 4.2.1, the administrator must first login by supplying the username specified here as “admin” and a password specified as “12345”.

After successfully supplying these information, the “Login” button is clicked. Once the “Login” is clicked, the proposed database driven web-based NYSC posting system checks and verify that the information provided are complete and correct. Otherwise, it clears all information entered and takes the administrator back to the page shown in Fig. 14 so as to refill the required information correctly.

On the other hand, if the information entered on the login form in Fig. 14 is correct, then clicking on the “Login” buttons automatically opens the posting page shown in Fig. 15 which was briefly discussed and shown in Fig. 5. The posting process as shown on the posting page in Fig. 15 is a simple three-step process which involves: 1) Select State, 2) Select School, and 3) POST. In continuation of the Benue State University as a case study, Benue is selected as the state from the “Select State” drop down menu as shown in Fig. 16. Next, Benue State University is selected from the “Select School” drop down menu. Immediately, the school is selected, Fig. 16 changes to Fig. 17 with the “POST” button included. Observe that all the three students of Benue State University that were registered are automatically displayed on the posting page. One can also observe that Akpan Vincent has already been posted to Akwa Ibom since his spouse resides in that state even though his call-up number is yet to be assigned.
Finally, clicking on the “POST” button automatically generates the call-up numbers and posts the prospective NYSC corp members to their respective states and regions for the NYSC service year. Note that the eight digit call-up number generated here is combination of the school’s authentication code (12362) and the serial number starting from 001 in the order in which the students were registered. However, the posting is done according to the criteria state in the objective and scope of this study and also shown in Fig. 5 and Fig. 15 under the caption “NOTE”.

4.3.2 Discussion of the Results

In the course of this study a database driven web-based NYSC posting system has been designed and its implementation strategy has also been demonstrated using the Benue State University as a typical case study. Some results from the implementation and posting of prospective NYSC corp members have been shown in Fig. 18. Before discussing the results, the three criteria for posting a prospective corp member are restated again as follows:

(a) Prospective corp members above the age of 30 years are exempted from the NYSC service programme;
(b) Prospective corp members who are below the age of 30 years and married are posted to the state where their spouse is located; and
(c) Prospective corp members below the age of 30 years and single are posted to a state and to a region different from their state of origin and state of residence.

In line with the criteria for posting of prospective corp member, it is evident from Fig. 18 where Akpan Vincent has been posted to Akwa Ibom State. Since he is below 30 years of age, he is qualified to participate in the NYSC programme for the year. Furthermore, even though he is from a neighboring Cross River State but because he is married and his spouse resides in Akwa Ibom State, it is justifiable for his posting to Akwa Ibom State.

In the second case, Bassey Victory is above 30 years of age and is automatically exempted from the NYSC programme. However, in the third case, Etuk Utomobong is single, she is below 30 years of age, and she is from the Enugu State which is in the Eastern region of the country. Therefore, her being posted to Kaduna State in the Northern region of the country is absolutely in line with the criteria for posting prospective NYSC corp member as specified in this study.

The details of how the activities of Fig. 6 through Fig. 18 are handled by the techniques and methodologies proposed in this study are summarized by Fig. A1 through Fig. A.4 given in Appendix A.

5. CONCLUSION

A database driven web-based national youth service corps (NYSC) posting system has been designed and implemented in this work. The design of the database driven web-based NYSC posting system has been accomplished with the aid of five major software tools, namely: 1) Apache web server for hosting the application, 2) Adobe Dreamweaver for writing HTML for the web-based application, 3) PHP for writing scripts required for the application, 4) MySQL as the database server for writing SQL, 5) Google Chrome browser for reading and displaying information developed using the first four tools. Significant instructions on the installation, setting-up and configuring these five software tools have also been presented which enables these tools to work interactively. In order to demonstrate the feasibility of the proposed database driven web-based application, an NYSC posting scheme has been used. In this NYSC posting scheme, prospective corp members are posted based on several criteria which are summarized into three groups.

Fig. 17: The form for posting NYSC corp members by the administrator after selecting the school.

Fig. 18: Posting or exemption of NYSC corp members their respective states and regions.
These are prospective corp members above the age of 30 years are exempted from the NYSC service programme; 2) prospective corp members who are below the age of 30 years and married are posted to the state where their spouse is located; and 3) prospective corp members below the age of 30 years and single are posted to a state and to a region different from their state of origin and state of residence. The results obtained from the implementation of the proposed database driven web-based NYSC posting system show the validity and efficiency of the NYSC posting system. It may be agreed that the proposed NYSC posting system can be deployed to NYSC headquarters for the posting of NYSC corp members to replace and speed up the manual method of posting NYSC corp members.

The proposed database driven web-based NYSC posting system has been designed and implemented for posting prospective NYSC corp members to different states in different regions based on some criteria at the national level. The proposed design methodology as well as the NYSC posting system can be adapted for posting corp members within a state to respective public and private sectors, industries and parastatals based on corp members’ course of study and areas of specializations. While the design methodology presented in this work has been formulated based on if and else statements; the development of mathematical models could be a good alternative for performance comparison in terms of execution speed and memory management. A more efficient approach could be the application of artificial adaptive neural fuzzy interference system to develop rule-based logic upon which posting could be made. It should be noted that for the two design techniques proposed and recommended here, the implementation platforms and software development tools is most likely to change from the five software development tools used in this workflow to design and implement the database driven web-based NYSC posting system.

REFERENCES


ABOUT THE AUTHORS

**Vincent A. Akpan** holds a Ph.D degree in Electrical & Computer Engineering from the Aristotle University of Thessaloniki (AUTH), Thessaloniki, Greece in 2011. He is currently a Lecturer I at FUTA, Akure, Nigeria. He has been an Visiting as well as an Adjunct Lecturer to a number of Universities in Nigeria with the Departments of Computer Science, Electrical and Electronics Engineering, Information & Communication Technology (ICT) and Mechatronics Engineering. His integrated research interests include: intelligent control, computational intelligence, instrumentation, real-time embedded systems, signal processing & machine vision, nonlinear system identification, and Robotics & Automation. He is the co-author of a book and has authored and/or co-authored more than 40 articles in refereed journals and conference proceedings. He is a regular reviewer in more than 20 local and international scientific/academic journals and several IEEE sponsored conferences. He is the managing editor of the Nigerian Journal of Pure and Applied Physics (NJPAP) as well as an editorial board member for the American Journal of Intelligent Systems (AJIS). Dr. Akpan is a member of the IEEE, USA, The IET, UK and a fellow of IPMD, Nigeria. He was one among the two Nigerian recipients of the 2005/2006 Greek State Scholarship (IKY) for a Ph.D. programme tenable in Greece.

**Rachael O. Vincent-Akpan** received her B.Sc. in Accounting from the University of Ilorin, Nigeria in 2001; a Master of Business Administration (MBA) from The Federal University of Technology, Akure, (FUTA) Nigeria in 2006; and a Postgraduate Diploma (PGD) in Computer Science, FUTA, Akure, Nigeria in 2013. Since 2003 she has been with the Bursary Department, FUTA, Akure, Nigeria where she is currently a Principal Accountant and has served in several Units of the Bursary Department in FUTA. Mrs. Vincent-Akpan is an Associate Member of the Institute of Chartered Accountants of Nigeria (ICAN).

**Olumide O. Obe** obtained B.Tech. degree in Computer Science from the Federal University of Technology, Akure, Nigeria in 2000; a M.Sc. in Computer Science from the University of Lagos, Akoka, Nigeria in 2002; and a Ph.D degree in Systems Engineering from the Polytechnica University of Bucharest, Romania in 2010. He was a Teaching and Research Assistant at the University of Lagos between 2001 and 2003. He has been a lecturer at the Computer Science Department of the Federal University of Technology, Akure since 2004. He was an adjunct lecturer to a number of Universities in Nigeria. His research interest area includes: Artificial Intelligence, hybrid adaptive intelligent control systems, Robotics & Automation, Mobile Applications, Security, and Big Data. He has co-authored a book and has authored/co-authored more than 30 articles in refereed journals and conference proceedings. He is a regular reviewer in local as well as international scientific/academic journals. Dr. Obe is a member of the IEEE, The ACM, and The IAENG, UK. He was a scholar of the Romanian Ministry of Education. He is a fellow of the Massachusetts Institutes of Technology, USA.

**Reginald A. O. Osakwe** received his Ph.D in electronics from the University of Ibadan, Nigeria in 1995. Since 2007 he has been with the Department of Physics, The Federal University of Petroleum Resources, Effurun, Delta State, Nigeria where he is currently a Senior Lecturer. During the periods 1998 – 2009 he has been a visiting lecturer to the Shell Intensive Training Programme (SITP), Warri, Delta State under the Shell Petroleum Development Company (SPDC) of Nigeria. His research interests are on microprocessor architecture, embedded computer systems, chaotic dynamics, and image and signal processing systems. Dr. Osakwe is a member of the Nigeria Institute of Physics (NIP). He was a recipient of The Federal Government of Nigeria Scholarship to study in the U.S.A. leading to the award of an M.Sc.
Table A.1: Universities with their state identification number (State_ID) and authentication number used for the study.

<table>
<thead>
<tr>
<th>S/N</th>
<th>School</th>
<th>State_ID</th>
<th>Selected</th>
<th>Authentication</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>-PLEASE SELECT UNIVERSITY-</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>Bingham University</td>
<td>2</td>
<td></td>
<td>12345</td>
</tr>
<tr>
<td>3.</td>
<td>University of Abuja</td>
<td>2</td>
<td></td>
<td>12346</td>
</tr>
<tr>
<td>4.</td>
<td>Bale University</td>
<td>2</td>
<td></td>
<td>12347</td>
</tr>
<tr>
<td>5.</td>
<td>Turkish Nile University</td>
<td>2</td>
<td></td>
<td>12348</td>
</tr>
<tr>
<td>6.</td>
<td>Abia State University</td>
<td>3</td>
<td></td>
<td>12349</td>
</tr>
<tr>
<td>7.</td>
<td>Abia Polytechnic</td>
<td>3</td>
<td></td>
<td>12350</td>
</tr>
<tr>
<td>8.</td>
<td>University of Adamawa</td>
<td>4</td>
<td></td>
<td>12351</td>
</tr>
<tr>
<td>9.</td>
<td>Adamawa Polytechnic</td>
<td>4</td>
<td></td>
<td>12352</td>
</tr>
<tr>
<td>10.</td>
<td>Akwa Ibom University</td>
<td>5</td>
<td></td>
<td>12353</td>
</tr>
<tr>
<td>11.</td>
<td>Akwa Ibom Polytechnic</td>
<td>5</td>
<td></td>
<td>12354</td>
</tr>
<tr>
<td>12.</td>
<td>Anambra University</td>
<td>6</td>
<td></td>
<td>12355</td>
</tr>
<tr>
<td>13.</td>
<td>Anambra Polytechnic</td>
<td>6</td>
<td></td>
<td>12356</td>
</tr>
<tr>
<td>14.</td>
<td>Bauchi State University</td>
<td>7</td>
<td></td>
<td>12357</td>
</tr>
<tr>
<td>15.</td>
<td>Federal University of Bauchi</td>
<td>7</td>
<td></td>
<td>12358</td>
</tr>
<tr>
<td>16.</td>
<td>Bayelsa State University</td>
<td>8</td>
<td></td>
<td>12359</td>
</tr>
<tr>
<td>17.</td>
<td>Federal University of Balyelsa</td>
<td>8</td>
<td></td>
<td>12360</td>
</tr>
<tr>
<td>18.</td>
<td>Benue State University</td>
<td>9</td>
<td></td>
<td>12361</td>
</tr>
<tr>
<td>19.</td>
<td>University of Agriculture</td>
<td>9</td>
<td></td>
<td>12362</td>
</tr>
<tr>
<td>20.</td>
<td>University of Maiduguri</td>
<td>10</td>
<td></td>
<td>12363</td>
</tr>
<tr>
<td>21.</td>
<td>Borno State University</td>
<td>10</td>
<td></td>
<td>12364</td>
</tr>
<tr>
<td>22.</td>
<td>University of Calabar</td>
<td>11</td>
<td></td>
<td>12365</td>
</tr>
<tr>
<td>23.</td>
<td>Cross State University</td>
<td>11</td>
<td></td>
<td>12366</td>
</tr>
<tr>
<td>24.</td>
<td>Delta State University</td>
<td>12</td>
<td></td>
<td>12367</td>
</tr>
<tr>
<td>25.</td>
<td>Bonyi State University</td>
<td>13</td>
<td></td>
<td>12368</td>
</tr>
<tr>
<td>26.</td>
<td>Edo State University</td>
<td>14</td>
<td></td>
<td>12369</td>
</tr>
<tr>
<td>27.</td>
<td>Federal University of Benin</td>
<td>14</td>
<td></td>
<td>12370</td>
</tr>
<tr>
<td>28.</td>
<td>Ekiti State University</td>
<td>15</td>
<td></td>
<td>12371</td>
</tr>
<tr>
<td>29.</td>
<td>University of Nsukka</td>
<td>16</td>
<td></td>
<td>12372</td>
</tr>
<tr>
<td>30.</td>
<td>Enugu State University</td>
<td>16</td>
<td></td>
<td>12373</td>
</tr>
<tr>
<td>31.</td>
<td>Imo State University</td>
<td>18</td>
<td></td>
<td>12374</td>
</tr>
<tr>
<td>32.</td>
<td>Federal University of Technology Owerri</td>
<td>18</td>
<td></td>
<td>12375</td>
</tr>
<tr>
<td>33.</td>
<td>Igawa State University</td>
<td>19</td>
<td></td>
<td>12376</td>
</tr>
<tr>
<td>34.</td>
<td>Kaduna State University</td>
<td>20</td>
<td></td>
<td>12377</td>
</tr>
<tr>
<td>35.</td>
<td>Federal University of Kaduna</td>
<td>20</td>
<td></td>
<td>12378</td>
</tr>
<tr>
<td>36.</td>
<td>Kano State University</td>
<td>21</td>
<td></td>
<td>12379</td>
</tr>
<tr>
<td>37.</td>
<td>Katsina State University</td>
<td>22</td>
<td></td>
<td>12380</td>
</tr>
<tr>
<td>38.</td>
<td>Kebbi State University</td>
<td>23</td>
<td></td>
<td>12381</td>
</tr>
<tr>
<td>39.</td>
<td>Kwara State University</td>
<td>24</td>
<td></td>
<td>12382</td>
</tr>
<tr>
<td>40.</td>
<td>Federal University of Ilorin</td>
<td>24</td>
<td></td>
<td>12383</td>
</tr>
<tr>
<td>41.</td>
<td>Ogun State University</td>
<td>25</td>
<td></td>
<td>12384</td>
</tr>
<tr>
<td>42.</td>
<td>Ondo State University</td>
<td>25</td>
<td></td>
<td>12385</td>
</tr>
<tr>
<td>43.</td>
<td>Ondo State University</td>
<td>26</td>
<td></td>
<td>12386</td>
</tr>
<tr>
<td>44.</td>
<td>Ondo State University</td>
<td>26</td>
<td></td>
<td>12387</td>
</tr>
<tr>
<td>45.</td>
<td>Osun State University</td>
<td>27</td>
<td></td>
<td>12388</td>
</tr>
<tr>
<td>46.</td>
<td>Osun State University</td>
<td>28</td>
<td></td>
<td>12389</td>
</tr>
<tr>
<td>47.</td>
<td>Ogun State University</td>
<td>28</td>
<td></td>
<td>12390</td>
</tr>
<tr>
<td>48.</td>
<td>Ogun State University</td>
<td>29</td>
<td></td>
<td>12391</td>
</tr>
<tr>
<td>49.</td>
<td>Osun State University</td>
<td>30</td>
<td></td>
<td>12392</td>
</tr>
<tr>
<td>50.</td>
<td>Osun State University</td>
<td>30</td>
<td></td>
<td>12393</td>
</tr>
<tr>
<td>51.</td>
<td>Osun State University</td>
<td>31</td>
<td></td>
<td>12394</td>
</tr>
<tr>
<td>52.</td>
<td>Osun State University</td>
<td>31</td>
<td></td>
<td>12395</td>
</tr>
<tr>
<td>53.</td>
<td>University of Ibadan</td>
<td>32</td>
<td></td>
<td>12396</td>
</tr>
<tr>
<td>54.</td>
<td>Oyo State University</td>
<td>32</td>
<td></td>
<td>12397</td>
</tr>
<tr>
<td>55.</td>
<td>University of Jos</td>
<td>33</td>
<td></td>
<td>12398</td>
</tr>
<tr>
<td>56.</td>
<td>Plateau State University</td>
<td>33</td>
<td></td>
<td>12399</td>
</tr>
<tr>
<td>57.</td>
<td>University of Port Harcourt</td>
<td>34</td>
<td></td>
<td>12400</td>
</tr>
<tr>
<td>58.</td>
<td>Rivers State University of Technology</td>
<td>34</td>
<td></td>
<td>12401</td>
</tr>
<tr>
<td>59.</td>
<td>Sokoto State University</td>
<td>35</td>
<td></td>
<td>12402</td>
</tr>
<tr>
<td>60.</td>
<td>Taraba State University</td>
<td>36</td>
<td></td>
<td>12403</td>
</tr>
<tr>
<td>61.</td>
<td>Kogi State University</td>
<td>37</td>
<td></td>
<td>12404</td>
</tr>
<tr>
<td>62.</td>
<td>Yobe State University</td>
<td>38</td>
<td></td>
<td>12405</td>
</tr>
<tr>
<td>63.</td>
<td>Zamfara State University</td>
<td>39</td>
<td></td>
<td>12406</td>
</tr>
</tbody>
</table>
Fig. A.1: The form for posting NYSC corp members based on course of study and departmental identification number.

Fig. A.2: The form for posting NYSC corp members based on university authentication with state and student’s identification numbers.
Fig. A.3: The form for posting NYSC corp members based on place, region and country of study.

Fig. A.4: The form for posting NYSC corp members based on state of origin, age and University with call-up and batch numbers.
Enhancement in Security by Reducing Dimensions of Hyperspectral Face Images for Face Recognition

S. Arya, N. Pratap & V. Kumar
Department of Computer Science
Gurukula Kangri Vishwavidyalaya, Haridwar
Neelkanth Institute of Technology, Meerut
neerajpr23@gmail.com
Phone: +919897335329

ABSTRACT
From the past few decades the security system industry has grown rapidly. Every year there is an increased demand for more precise, accurate and full proof security systems. Designing flawless advance security system involves complex parameters and the most important one is the image. This has attracted researchers in the field of Face Recognition (FR). Due to the high dimension of the images, it’s always a challenging task to collect them and analyze them. Generally the depth of dimension is the main problem, especially in the case of hyperspectral face images, decreases the classification accuracy thus affecting the performance of a security system. During experimentation, it has been observed that reduction in dimension results in enhancing classification accuracy as well as recognition accuracy, so directly affect the security performance. The objectives of this paper are:
1) Study and review the major Data reduction techniques along with their advantages and disadvantages.
2) Relevance of Data Reduction techniques in improving the overall accuracy of FR System and
3) Implementation of FR systems in designing advance security system.

Keywords- PCA; Face Recognition; Preprocessing.

I. INTRODUCTION

Generally in a conventional face recognition system, broadband monochrome or color imaging sensors [17], only those imagery data can be seen and analyzed which include only visible light in Electromagnetic Spectrum (EMS). The CBI contains the huge difference between the low visual features and high level and is pointed by the color, texture and shape. Due to the low-level visibility features, the representation of these images is not possible directly by the human. The example of these types of images are multispectral images, hyperspectral images [16], images related to medical science etc and contains thousands of features to represent an image. When the computer deals all these types of images, all these large volume of features produce problem in analysis. This decrease the performance as well as efficiency of the system. So it is required to reduce the dimension of these images without the loss of information. An example of hyperspectral and multispectral images are shown with Figure 1.

Dimensionality reduction is the process of converting the high dimensional patterns into the low dimensional pattern and also to maintain the local structure of the original high dimensional data. Dimensionality reduction is the process to remove all the redundant features and to extracts relevant features.
Intrinsic dimensionality is the term; specify the required parameters to point the properties of the data. So it is always required that the reduced representation should correspond to the intrinsic dimensionality. The main advantage of lower dimensional representation is the efficient visualization of the data and in the same way the high dimensional data can be analyzed by transforming them into a lower dimensional representation. For example, the hyper spectral satellites produce the images containing thousands of bands and these large set of bands creates problem in analysis of these images. So it is always required to reduce these bands and by selecting those, which produce the relevant information. Generally there are two techniques related with the reduction technique i.e. Linear and non-linear techniques. In linear, the large dimensional data set is transformed into low dimensional data set by using the linear mapping. In other words if m is the original high dimensional data set and n is the lower dimensional data set, then n is always less than to m.

Linear techniques mainly focus on two methods, one is Principal Component Analysis (PCA) [3] and another is Linear Discriminant Analysis (LDA) [4]. The main advantages of linear techniques are that these are very simple and always easy to implement. In some situation the linear techniques are limited to find the accurate structure of non-linear datasets. These can be overcome by introducing the non-linear techniques such as Locally Linear Embedding (LLE) [5] and Isomap [6].

The main advantage of non-linear methods is the capability to handle complex non-linear structures of large datasets in the efficient manner. Sometime there can be the situation that the required data in the high dimensional data exist on the abstract space, called as manifold. This type of non-linear reduction methods are called as manifold learning techniques [5][6]. In Manifold learning, the non-linear structures are extracted from high dimensional data sets and maintain the same structure as in the high dimension. In such cases, the linear methods are limited while the non-linear dimensionality reduction techniques can find the manifold structure efficiently.

2. FACE IMAGE CLASSIFICATION

Image classification is the process of sorting pixels into finite number of individual classes or categories of data, based on their DN values of pixels. A classified image is comprised of an array of pixels, each of which belong to a particular theme and called thematic image. Classification of multi-dimensional images is used to assign corresponding level with respect to groups of same characteristics, with the aim of discriminating multiple objects from each other with in the image(s). The main objective of image classification is to establish spatial relationship with neighbouring pixels and assign all pixels in the image to a particular or different class(s) [20]. Image classification process is categorized into two forms:

Supervised classification- with supervised classification, we identify similar area on an image by identifying ‘training sites’, also known Area of Interest (AOI), of known targets or classes then extrapolating those to other area of unknown targets. There are different supervised classification algorithms in DIP research domain named as, Quadratic Discriminant Analysis (QDA), Minimum Distance, Parallel Piped, Maximum likelihood [20], Linear Discriminant Analysis (LDA), Support Vector Machine [19].

Unsupervised classification- with unsupervised classification, the user does not require specifying any ‘training sites’ or AOI for classification. The classification creates natural grouping of pixels in the image, called clusters or classes. There are different unsupervised classification algorithms named as c-Means clustering, Active Classification through Clustering (ACC), K-means clustering etc [18]. In today scenario, the main objective of image processing is subject visualization, enhancement, sharpening and restoration, transformation, classification as well as measurement of patterns in different areas of research and applications. The Face Recognition is a technique of the pattern matching techniques which is used to identify and verify the individual human face.

3. FACE IMAGE EXTRACTION SYSTEM

As the FR has wide applications in the different fields such as visual surveillance, retrieval of an identity from a face database for forensic applications, criminal investigations and security systems, so it has gain the more attention by the researchers to enhance the efficiency. Face recognition also face the dimension reduction problem because for calculation, if a face image having p*q pixels can be represented by a vector in Tp*q. This p*q-dimensional space is too high for the recognition system and increase the complexity for recognition process. So it is always required to reduce the dimension of data sets so that’s the fewer variables are used for the computation purpose.

In relation to the machine learning, image extraction can be seen as the problem of dividing the images into different classes and the accuracy is represented by performance metric. To measure the performance of machine learning system, the classification accuracy is a widely used metric. Generally a FR system contains two steps, one is Feature Extraction and other is Classification. In feature extraction, the features are extracted for the representation of faces and belong to a particular class. The feature extraction is performed with low dimensional data sets, obtained by linear or non-linear dimensionality reduction methods. Then the classifier is trained for training images to classify the face images into the appropriate classes.
4 LITERATURE REVIEW

A number of studies have been performed in the field of face detection and recognition. Generally the Face recognition methods are based on two techniques named as feature based methods and subspace methods. In feature – based methods the local features of the face such as the position of the eyes, nose, mouth etc are collected while on the other hand the Subspace method works on the reduction of the dimension of the data and maintain the maximum distinction between distinct classes.

The subspace methods works on the two approaches named as ‘Eigenface’ [13] and ‘Fisherface’ [4]. In ‘Eigenface’ PCA is used as the linear unsupervised dimensionality reduction method to generate the subspace while the ‘Fisherface’ works on Linear Discriminant Analysis (LDA) as the linear supervised dimensionality reduction method.

PCA finds a projection on a lower dimensional representation and provides the correct representation of the data with minimum error and also finds the best axis for projection [4]. On the other hand LDA maximize the separation between different classes with minimum distances among them. In classification systems LDA is given too much preference because it generates the high class discrimination by using the class information [3] [4].

In 2001, Aleix M conducted a study in which PCA performs better than to LDA, when the number of samples per class is very less.

LLE and Isomap are used for FR [5] [11] and are based on the non-linear dimensionality reduction techniques. The high dimensional data that contains non-linear structures is processed by LLE and also maintain neighbourhood structure of the high dimensional data. LLE show the less discrimination capability as compared to LDA [11]. In 2000, Joshua B [6] conducted a study by using Isomap to maintain the true dimensionality and geometric structure of the high dimensional data. The result was in favour as it maintains the global structure but the computational cost was too high for large datasets. It was also experimentally proved that when Isomap is applied alone for face recognition, the performance was satisfactory [11].

Several another studies was also conducted in which the different dimension reduction techniques were combined for face recognition [8] [12] However, the performance achieved was not so high.

In 2010, Mahesh Pal et al. [7] conducted a study on the dimension reduction as a pre-processing step. Support Vector Machine (SVM) is a widely used method for classification which contains a hyper plane, which provides the high distance between different classes to provide better classification accuracy. It was experimentally proved that the classification accuracy of SVM can be increased by reducing the dimensionality of the data [7]. As the features are increased, the classification accuracy of SVM decreases. So SVM is best apparent for small training sets. Experimental results proved that there is a strong dependency between dimensionality reduction and classification accuracy of SVM. When the feature dimensions were taken 55, 60 and 65, the accuracy achieved was 92.24%, 92.11% and 91.76%.

Besides providing classification accuracy, dimensionality reduction offers the advantages to SVM such as increment in the speed of the classification by reducing feature set size and also reduced in data storage requirements. In 2009, Jianke Li et al. [8] conducted a study and introduced a method for FR based on the combination of Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) to extract the feature. In order to perform feature extraction and dimension reduction, PCA is used, resulting that the components are uncorrelated and have less reconstruction error. After that, LDA is used to increases the class separation and provides higher discrimination of the samples. So PCA makes a PCA subspace and the combination of PCA and LDA form the LDA subspace showing high discrimination.

To perform the classification, Nearest Neighbor Classifier (NNC) was used on the ORL face database containing 400 images of 40 individuals. The accuracy was calculated for different training sample size and for various feature dimensions. As the sample size was increased, the classification accuracy of both PCA method and combined PCA and LDA method was also increased. Also, it is observed that the combination of PCA and LDA provides better classification accuracy than the PCA alone method for face recognition [8].

In 2009, Md. Omar Faruque et al [9] proposed a face recognition system using Principal Component Analysis (PCA) and support vector machine (SVM). Because the major components of a FR system are feature extraction and classification, so it’s always a matter of consideration, how the features are extracted and how the features are classified into the several classes. Therefore, the selection of the feature extractor and selection of classifier is very important in a face recognition system. In this study, the authors have used PCA for feature extraction and SVM for classification.
The classification based on SVM is done by producing the new hyper plane which provides maximum separation between the data item. The vectors near to the hyper plane are termed as support vectors. Input space point is non-linearly mapped into a high dimensional feature space by using kernel functions. In 2009, Md. Omar Faruque [9] used the kernel functions and conducted an experiment on the ORL face database containing 400 images of 40 individuals. They selected 200 samples as the training sample, used for constructing Eigen faces and also for training the SVM. From the experiments it can be observed that the combination of PCA and SVM [15] provides better classification accuracy for face recognition on the ORL face database.

In 2010, Sangwoo Moon et al [10] works on the dimensionality reduction technique using support vector machine, in which the decision vectors were used for classification purpose in SVM as the mapping vectors for dimensionality reduction and produces a feature set having high efficiency and better classification ability. Generally the nature of these mapping vectors are highly redundant. So, it is required to further reduce the redundancy, based on the factor called as asymmetric decorrelation and removes the less meaningful mapping vectors. The remaining mapping vectors are far away from each other and also provide higher classification accuracy. Experiment was performed on the handwritten numeric characters dataset and the classification accuracy was 98.2%. So it can be stated on the above facts that SVM provides better dimensionality reduction as well as higher classification accuracy.

Table 1 provides the brief summary of the different dimensionality reduction methods used for face recognition. Experiments were performed on the ORL database containing 400 images of 40 individuals and the face images were taken at different times under varying lighting and different facial expressions having the dark homogeneous background. It can be concluded from the literature survey, whatever be the classifier used for classification; dimensionality reduction is always a required step prior to the classification stage [7]. The classification provides higher classification accuracy preceded by the better dimensionality reduction stage. When the Nearest Neighbor Classifier (NNC) is used on face images in ORL database, PCA provided a classification accuracy of 85%, and significantly improved upto 93% by using PCA and LDA combination. So it can be stated that performance of combination of PCA and LDA is very much significant over to the PCA alone for improving classification accuracy of SVM [8].

Table 1: Summary of Classification accuracy corresponding to different methods

<table>
<thead>
<tr>
<th>Methods used</th>
<th>Selected Classifier</th>
<th>Dimensions of Features</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Principal Component Analysis</td>
<td>Nearest Neighborhood Classifier</td>
<td>40</td>
<td>85% Approx.</td>
</tr>
<tr>
<td>Principal Component Analysis</td>
<td>Support Vector Machine</td>
<td>40</td>
<td>97% Approx</td>
</tr>
<tr>
<td>Combination of Principal component analysis and Linear Discriminant Analysis</td>
<td>Nearest Neighborhood Classifier</td>
<td>40</td>
<td>93% Approx</td>
</tr>
</tbody>
</table>

It was also proved experimentally conducted on the ORL face database using SVM as a classifier, it is clear that PCA alone method provides better classification accuracy of 97% for the same feature dimension [9].

5. CONCLUSION AND FUTURE WORK

The feature dimensionality has great impact on the classification accuracy of a classifier. SVM is a widely used classifier, but prior to SVM, dimension reduction is mandatory step. PCA is one of the best techniques for dimension reduction and found suitable for SVM. By providing best dimension reduction and higher class discrimination prior to the classification, resulting higher classification accuracy. PCA is an extensively used dimensionality reduction method but limited in class discrimination. LDA improves the classification accuracy by introducing between PCA and SVM. Combination of PCA and LDA improved the capability of LDA when the samples of images are less and classification accuracy is also increased. During the dimension reduction using PCA, main features of face images are extracted, LDA selects the significant features for class separability and SVM classifier that classify them into one of available classes. As the Classification accuracy is increased.

FR accuracy percentages are also increased and also have a direct enhancement in security system. Security is directly depends on the accuracy of FR and FR accuracy is increased as the classification accuracy is increased. Future work also needs to explore new techniques to remove the redundant mapping vectors of SVM to minimize the computation cost and overall performance of the system.
REFERENCES


Multi-agent Based Course Allocator Using GAIA Methodology and JADE Framework

K.T. Igulu
Department of Computer Science
Rivers State Polytechnic, Bori
Rivers State, Nigeria.
E-mail: igulukt@gmail.com,
Phone: +2347018827522

Z.P. Piah
Department of Computer Science
Rivers State Polytechnic, Bori
Rivers State, Nigeria.

P.O. Asagba PhD
Department of Computer Science
University of Port Harcourt
Rivers State, Nigeria.

ABSTRACT

This paper discusses our work on the development of an interactive multi-agent system that automates university course allocation. The system receives as input the courses to be offered in a particular semester, their respective credit units and appropriately allocate it to competent lecturers based on their teaching profile. The system is composed of two agent types- The administrator agent and the lecturer agents which were the final concrete artifacts of the system. Four roles (allocation handler, environment monitor, administrator assistant and lecturer assistant) were discovered in the analysis stage. However, the allocation handler and environment monitor were subsumed by the administrator and lecturer agents at the design stage. The elicitation, analysis, design to implementation were quite natural. This proves Agent-Based Software Engineering (ABSE) as a viable paradigm. We used GAIA methodology for analysis and design and Java Agent Development (JADE) framework for implementation.

Keywords- Multi-agent; course allocation; Gaia, JADE; software engineering; intelligent, agent.

I. INTRODUCTION

It is a known fact that scheduling is an NP problem. A problem where there is no universally accepted optimal solution. Most Universities do this manually and of course becomes so tedious with the growing number of students and courses offered in a University. Along with the innate limitations or problems of traditional manual systems, manual course allocation and scheduling has the following key problems: 1) Keep and manage record of the previous data. 2) Meeting person. 3) Manage multiple queries for the same subject. 4) Make availability of the interested subjects for a faculty member. 5) Manage rooms for delivering lectures. Every university has a number of schools with respective heads and each school has a number of departments with respective heads. A department is composed of lecturers, students and other teaching enhancement facilities.

In a well-established university, a department should offer degrees ranging from undergraduate to post graduate. Each degree depending on the duration, has several levels (e.g., Bachelor of Science in Computer Science should have first, second, third and fourth year level students) administered concurrently in every academic session. The task of allocation is usually the job of the head of the department or can also be delegated to another fellow who proves competent. Allocation of this kind is never void of irregularities and anomalies. The problem faced by academic departments is the inefficient/ inappropriate allocation of courses to competent lecturers. In our interview with the faculty in charge of course allocation in the department of Computer Science, University of PortHarcourt, the following were discussed; 1.) Allocation is dedicated to a particular lecturer 2.) Number of courses to be offered is known a priori, 3.) Number of classes is also known a priori, 4.) The faculty needs information regarding the areas of expertise of the lecturers. 5.)
The subjects are communicated to lecturers and they are required to give their choices according to their levels and a minimum number of credit units to offer. 6.) When reply is received, courses-lecturers table is made. 7.) The course allocation administrator willfully allocates the courses. 8.) The number of courses to be allocated is determined by the designation of the lecturer. Professors are given less but highly technical load. 9.) Designation also determines the level course that can be assigned to a lecturer. E.g: an MSc holder can’t be allowed to teach doctorate students. 10.) Conflict is bound to occur. 11.) These conflicts are usually resolved on a one to one basis. 12.) The administrator lecturer prepares an allocation chart. 13.) Conflict is bound to occur.

This work discusses a typical university course allocation problem using ABSE paradigm. The system receives as input the courses to be offered in a particular semester, their respective credit units and appropriately allocate it to competent lecturers based on their teaching profile. The system is composed of two agent types- The administrator agent and the lecturer agents which were the final concrete artifacts of the system. Four roles (allocation handler, environment monitor, administrator assistant and lecturer assistant) were discovered in the analysis stage. However, the allocation handler and environment monitor were subsumed by the administrator and lecturer agents at the design stage. Figure 1 shows the interactions among the agents in the system. There are m number of Lecturer agent instances and one instance of Admin agent. The Admin Agent performs the allocation, monitors the environment for any request and does resolve conflicts among lecturer agents.

Figure 1. Interactions in the system

In this work, we assume that 1.) Each level(e.g., first year) has a single section. 2.) No external lecturer is needed 3.) All courses are offered by the assumed department.

The paper is organized as follows: In section 2, we discuss works that are related to our work. In section 3, we discuss our motivations for agency and the general benefits of using ABSE paradigm. Section 4 introduces the Gaia methodology. Section 5 discusses the analysis phase of the system with respect to Gaia standards. Section 6 provides information about the design phase of the system. Section 7 gives the implementation details of the system. Finally, section 8 wraps up our work.

2. RELATED WORKS

In [17], the authors argued that for certain classes of problem, adopting a multi-agent approach to system development affords software engineers a number of significant advantages over contemporary methods. If a problem domain is particularly complex, large or unpredictable, it might be only way it can be reasonably addressed is to develop a set of modular components that are specialized at solving a particular aspect of it.
In [22] the system is composed of intelligent agents but does not follow any Agent-Based software Engineering (ABSE) conventions. Other notable works can be found in [5] and [6] that uses stimulated annealing and genetic algorithm respectively. [1] uses Heuristic approach in generating the schedule. Literatures on time table scheduling can also be found in [14,17,31,40,41,43] whereas works on Gaia and Jade can be found in [8,9,16,18,20,22,34]. It is worthwhile to note that most of the works focused on time tabling not course allocation.

3. MOTIVATION/BENEFITS OF AGENCY

The human-like characteristics of agents provide a high abstraction level which may simplify the modeling and implementation of systems for complex domains. Agents can be trusted to pursue their goals and take initiative to interact only when needed; this independence reduces the need for external communications. Their autonomy leads to encapsulation of functionality, and coupling is reduced because agents do not provide any control point to external entities. The following list highlights some of the main dimensions along which agent systems are believed to enhance performance, these aspects are further elaborated on in [19]:

**Computational efficiency** because concurrency of computation is exploited. This requires that the communication is kept minimal, e.g. by transmitting high-level information and results rather than low-level data.

**Reliability** Components that fail can be gracefully recovered. Agents with redundant capabilities or appropriate inter-agent coordination are found dynamically and can take up responsibilities of agents that fail.

**Maintainability** A system composed of multiple components is easier to maintain because of its modularity.

**Responsiveness** The modularity of a multi-agent system leads to the possibility of handling anomalies locally without propagating them to the whole system.

**Flexibility** Agents with different abilities can adaptively organize to solve a given problem. An agent can also have a number of plans for reaching its goal and adapt its strategy to changes in the environment.

In order to deal with the complexities, the timeliness response and avoiding subjective impositions of course allocation and yet maintaining robustness and flexibility, great level of autonomy must be maintained. We identified some conflicts that may arise in course allocation. These conflicts require negotiation without much influence on the negotiating parties.

In [24,27] intelligent software agents(or intelligent agents or simply agents) are characterized as being autonomous, proactive, reactive, social, flexible and robust, as well as situated in an environment which they can sense and act upon. We found AOSE ideal to deal with the problems identified above. In fact, Multi agent systems provide the modularity that we want and the agents’ social ability makes them capable of meeting different restrictions and goals through negotiation and collaboration.

4. GAIA METHODOLOGY

The Gaia methodology deals with both the macro-level (societal) and the micro-level (agent) aspects of systems [26]. The Gaia methodology is based on organizational metaphor. It is assumed that the software entity is a collection of various self-functioning roles of the organization with their dedicated responsibilities towards meeting the global organizational objective. In fact in most complex systems, the system could be modeled as composing of sets of organizations with their respective organizational responsibilities (functionalities).

A software system is conceived as the computational instantiation of a (possibly open) group of interacting and autonomous individuals (agents). Each agent can be seen as playing one or more specific roles: it has a well-defined set of responsibilities or sub goals in the context of the overall system and is responsible for pursuing these autonomously. Such sub goals may be both altruistic (to contribute to a global application goal) or opportunistic (for an agent to pursue its own interests). Interactions are no longer merely an expression of interdependencies, and are rather seen as a means for an agent to accomplish its role in the system. Therefore, interactions are clearly identified and localized in the definition of the role itself, and they help characterize the overall structure of the organization and the position of the agent in it [26].

The evolution of the activities in the organization, deriving from the autonomous execution of agents and from their interactions, determines the achievement of the application goal, whether an a priori identified global goal (as, e.g., in a workflow management systems where altruistic agents contribute to the achievement of a specific cooperative project), or a goal related to the satisfaction of individual goals (as, for example, in agent-mediated auctions, whose purpose is to satisfy the needs of buyer and seller agents), or both (as, for example, in network enterprises exploiting market mechanisms to improve efficiency)[26].
The organizational perspective leads to a general architectural characterization of Multi-agent System (MAS) as depicted in Figure 2. A simpler system will eventually be modeled as a single organization. However, as complexity increases, programming practices like modularity and encapsulation suggest decomposing the system into different sub organizations consisting of a considerable number of agents to pursue its (sub)-organizational goal(s). In an organization, agents may interact to share computational tasks or for exchange of knowledge. In such system as depicted in Figure 2, there could be intra-organization (within) or inter-organizational (outside) interactions. In some other complex computation, agents may need to migrate from one organization to another. An agent may perform one or more roles in an organization [9].

Moreover, the MAS is completely immersed in an environment which is basically an ensemble of resources that the agents need to interact with to accomplish their role. Of course, interaction with the environment occurs via some sort of sensors and actuators-mechanisms enabling agents to perceive and act upon some part of the environment. Such portion of visibility is determined by agent’s specific role, as well as by its current status.

The first proposed Gaia methodology consists of two iterative phases, analysis and design. Gaia does not address requirement elicitation but does not necessarily ignore it. Gaia is usually open to any platform of implementation but experience shows that it is better implemented with a FIPA compliant platform like JADE [16,34,35]. Thus our choice of implementation platform.

5. ANALYSIS PHASE

The analysis phase involves building conceptual or abstract models that may not directly impact the system. It assumes that the analyst has conceptualized the problem and is clear of what the system should and should not do. Here, roles are identified and their interactions are modeled. The two artifacts produced at the end of the analysis phase of Gaia are role model and interaction models. The roles may not be detailed at the analysis stage [9].

5.1 Role Models

Roles consist of four attributes- responsibilities, permissions, activities and protocols.

Responsibilities (Rs) are said to be a key attribute of a role since they determine the functionality. Responsibilities are of two kinds- liveness properties-the role has to add something good to the system and safety properties-the role must prevent that something bad happens to the system.

Permissions (Ps) are the rights associated with a role. They identify the resources that are available to that role in order to realize its responsibilities.

Activities of a role are computations associated with the role that may be carried out by the agent without interacting with other agents. Activities are thus “private” actions.

Protocols are computations that require interaction with other agents. (P, A) will be used for Protocols and Activities and (D) for description in the role model because of space.
With thorough analysis, we identified four roles in the system. Environment monitor, Lecturer Assistant Administrator Assistant and Allocation Handler.

Having identified the roles, we now move on to documentation of the roles. We will only show documentation of three roles-Administrator Assistant, Lecturer assistant and allocation Handler roles.

**Role Schema:** AdminAssistant[ADA]

- **D**
  - It publishes requests for bidding of courses; it subscribes to the yellow pages knowledgebase and receives notification if a LA role registers; It receives commitment messages from LAs and relays same to AH; it publicizes the result of allocation to respective LA.

- **P, A**
  - RegisterDF, QueryDF, SubscribeToDF, SaveLAs, RegisterCourse, UpdateCourse, BroadcastMessageForbidding, ReceiveCommitmentToOffer, SendAllocationInstruction, AllocationChartPublishing

- **Ps**
  - reads, updates LADataS, AllocationChart, CourseDataS, Creates LADataS, CourseDataS

- **Rs**
  - Liveness: RegisterDF. InitializeLADataS|| InitializeCourseDataS. SubscribeToDF. ((ReceiveMessageFromDF.SaveLAServices) ++||RegisterCourses. (SendRequestForAllocation.ReceiveCommitmentToOffer)+. PublishAllocationResult+)||[UpdateCourses||UpdateLA]*.[AlterAllocation]*

- **Safety:** CourseList not nil, LAlist not Nil, No Course is Allocated to more than one LA.

**Role Schema:** LecturerAssistant[LA]

- **Ds**
  - It acts on behalf of a lecturer; it does most intelligent decisions a lecturer is supposed to take; it registers its service(lecturer)to DF; it replies the ADA when sent message to decide course to offer by choosing from experience profile; the lecturer can also set the priority of course to choose.

- **P, A**
  - RegisterDF, DecideCoursePriority, ChooseCourse, QueryAvailableCourses, SaveAllocation, SendsSatisfactoryMessage

- **Ps**
  - Creates, reads, and updates own experience profile.

- **Rs**
  - Liveness: RegisterDF. InitializeCourseTaughtDataS, [QueryAvailableCourses]* ReceiveMessageToOfferCourse +. SendCommitmentToOffer+. ReceiveCourseAllocated+. SendsSatisfactoryMessage.SaveAllocation Safety: true

**Role Schema:** AllocationHandler[AH]

- **D**
  - Does the allocation computation; it resolve allocation conflicts when they arise, it forwards allocation chart to ADA

- **P, A**
  - AllocationOfCoursesToLecturers, CheckForConflict, SendAllocationStatusMessage

- **Ps**
  - Creates, reads and updates AllocationChart

- **Rs**

---

**Figure 4. Role and its attributes [26]**

**Figure 5. Admin Assistant Role model**

**Figure 6. Lecturer Assistant Role model**

**Figure 7. Allocation Handler Role model**
7.2 Interaction Model
The model consists of set of protocol definitions, one for each type of inter-role interaction. Here, a protocol can be viewed as an institutionalized pattern of interaction [9,26].

A protocol definition consists of the following:
- **Purpose**: Brief textual description of the nature of the interaction (e.g., “AllocateCourses”);
- **Initiator**: the role(s) responsible for starting the interaction (e.g. admin);
- **Responder**: the role(s) with which the initiator interacts;
- **Inputs**: Information used by the role initiator while enacting the protocol (list of courses to be allocated);
- **Output**: Information supplied by/to the protocol responder during the course of the interaction (e.g., allocation chart);
- **Processing**: brief textual description of any processing the protocol initiator performs during the course of computation.

We present some of the protocols of our system.

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Initiator</th>
<th>Responder</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PublishRequestForBidding</td>
<td>ADA</td>
<td>LA</td>
<td>Packages courses and sends request to all discovered LAs for bidding.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CommitmentToOfferMessage</td>
</tr>
<tr>
<td>UpdateAgentRecord</td>
<td>DFA</td>
<td>ADA</td>
<td>Yellow page notification</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Updated AgentRecord</td>
</tr>
<tr>
<td>SendAllocationMessage</td>
<td>ADA</td>
<td>AH</td>
<td>After Receipt of commitment to offer message from LA, ADA relays same to AH for allocation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CommitmentToOfferMessage</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Partial Allocation Chart</td>
</tr>
<tr>
<td>AllocationStatusMessage</td>
<td>ADA</td>
<td>LA</td>
<td>After Receipt of allocation status from AH, ADA sends the status to various LAs.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Partial Allocation Chart</td>
</tr>
</tbody>
</table>

Figure 8. Protocols related to Admin Assistant Role
6. SYSTEM DESIGN PHASE

The analysis phase is basically the conceptualization of the system. It produces the input for the design phase. The activities of design phase involve the transformation of the abstract entities (represented in role and interaction models) of the analysis phase to concrete entities that may have direct impact on the realization of the system[9,26]. Agent Type, Services and Acquaintances models are identified and documented in this phase. The succeeding sections present the artifacts of Gaia design phase with respect to our case study. We exclude discussion of the Service model for privacy reasons.

6.1 Agent Type Model

Agent types are the counterparts of objects in object-oriented approaches. They are basic design units of an agent-based system and their realization at runtime is agent instances. Agent types in the system under development are defined on the basis of the roles that they play. In most cases, there is one-one mapping between roles to agent types. Gaia represents Agent Type with a rectangle and a role with an oval [8]. It uses annotation to represent the number instances of such agent mapped and a directed edge from role to agent.

In our system, the Lecturer Assistant Role is mapped to Lecturer Agent whereas the Environment monitor fades out during analysis and allocation handler is subsumed by the Admin Agent. The agent type model is depicted in Figure 11 below.

6.2 Acquaintance Model

The Acquaintance model depicts the communication links existing between agent types. It is in fact a directed graph in which nodes represent agent types and arcs show communication pathways.

7. IMPLEMENTATION

As we have stated, we use JADE for implementation. We now discuss the implementation of the system based on JADE. JADE is FIPA compliant [16, 34, 35].

7.1 Admin Agent Class

The adminAgent class is the core class of this MAS. It embeds the core functionalities such as
- searching and discovery of lecturer agents,
- initiating conversation by sending appropriate messages to Lecturer Agents,
• Checking the appropriating of accepting a proposal from a lecturer agent
• Appropriately allocating courses to lecturer and
• Resolution of conflicts.

In order to implement Jade agents in GUI, Jade distribution comes with middleware support based on java Window tool Kit and Swing APIs. Our AdminAgent class extends the AgentGui instead of jade.core.Agent. In Jade, the tasks an agent performs are embedded in its behaviours. Usually, an agent has one or more behaviours depending. Jade comes with a number of behavioral classes which can be used to achieve diverse tasks of an agent[8,16, 44].

The following are the behaviours of Admin Agent;
- **OneShotBehaviour** is extended to achieve DF() registration.
- **TickerBehaviour** is used when the start allocation is triggered. This behavior searches the DF (Directory Facilitator[16]) every 60 seconds for 5 minutes after it has received the trigger. A TickerBehaviour repeats periodically using the millisecond parameter passed to its constructor. It does not stop except stopped explicitly. This behavior embeds the task of searching the DF for LECTURER_TYPE service of lecturer agents. It then adds the sendCallForProposalToAllLecturerAgent behavior that sends them CALL FOR PROPOSAL (CFP) (FIPA performative [10,12,13]) message with appropriate content. Here the content is a list of AgentAction objects called Offer (a class with Course attribute), a part of CourseAllocationOntology. The behavior also handles the incoming proposal and the rest of the conversation between lecturer agents and carries the allocation process.

- **WakerBehaviour** is used to terminate the Ticker behaviour by calling the stop() method of TickerBehaviour. This behavior waits for 6 minutes and then triggers the transfer of the content of partialAllocationTable to mainAllocationTable and also alerts the AgentGui.
- **SequentialBehaviour** embeds the behaviours described above.

7.2 Lecturer Agent Class
The Lecturer Agent (LA hereafter) handles all communication with Admin Agent using appropriate message attributes. It has two behaviours-OneShot and cyclic behaviours to register and communicate with Admin respectively.

7.3 Ontology
As human, we communicate with a using the symbol, syntaxes and semantic of a particular language. The language’s syntax and semantic must be known and understood by the communicating parties[7,32]. One problem that may occur in agent approach is the choice of format of encoding the content of a message (here, ACL messages).

Sending information as primitives (Strings, numbers or characters) will only turn to be efficient if we are interested in primitive values. Serializing objects would have been the next option but serialized object can’t be decoded on transit. Conveniently, Jade provides us with content language and ontology for converting and checking the semantics information encoded by another agent. One of the importance of ontology is that agent architecture may differ in the sense language and platform[7, 10,11,12,13,16].

It is clear however that, if on the one hand information handling inside an agent is eased, on the other hand each time agent A sends a piece of information I to agent B

- “A” needs to convert its internal representation of I into the corresponding ACL content expression representation and B needs to perform the opposite conversion.
- Moreover B should also perform a number of semantic checks to verify that I is a meaningful piece of information, i.e. that it complies with the rules of the ontology by means of which both A and B ascribe a proper meaning to I.

The support for content languages and ontologies provided by JADE is designed to automatically perform all the above conversion and check operations.

By Jade design recommendations, we found the structure suitable;
- CourseAllocationOntology
- CourseAllocationVocabulary
- Course as Concept
- Lecturer as Concept
- Problem as Concept
- Conflict as Concept
- Offer as Action
- Allocated as Predicate
- Taught as predicate

We adopted the SL0 as our content language.

7.4 Allocation Core and Conflict Resolution
Here we present how the allocation core is done and how the agent resolves several conflicts that could arise. For simplicity, we will depict the conflict resolution with a flowchart.
Figure 13. Conflict Resolution
8. CONCLUSION/ FUTURE WORK

We acknowledge the fact that considerable efforts have been channeled towards school timetabling research but it is worthy to note that few of the works uses the concept of agency and very few follows ABSE paradigm.

Our work- Multi-Agent Based Course Allocator is fully ABSE compliant. We follow the Gaia methodology convention in the analysis and design of the system. The final artifacts of the design phase (Agent model, service model and acquaintance model) are directly mapped to the various classes provided by JADE framework.

The Admin Agent performs the allocation, monitors the environment for any request and does resolve conflicts among lecturer agents. This system has proven to increase the efficiency and a high level of appropriateness of course allocation with respect to a lecturer’s choice and experience. The administrator can alter allocation when necessary and be able to update lecturer and course records when necessary. This is an extreme situation. As a matter of fact, the human administrator has less or no work except for this extreme case.

With respect to applications, ABSE is still at its infancy. In fact the software industry is yet to embrace it but it is envisaged that ABSE will be what OO (object-oriented) paradigm is today in less than two decades.

As future work,

- Adding learning capability. Learning in agent systems is a very interesting topic which we would have loved to have the time to investigate further. This capability could be introduced at various places in our design and would, at least in theory, which could eventuate to a more dynamic agent system.
- Making it a web-based so that system communicates with human via email.

ACKNOWLEDGEMENT

We acknowledge the researchers whose works were cited in our work especially the Gaia and JADE teams.
REFERENCES


An Investigation into the effect of channel estimation and interpolation techniques in long term evaluation system through realistic channel conditions

M. Setia & J. Malhotra PhD
ECE Department
GNDU-RC, India
Emails: manoramasetia@gmail.com, jyootesh@gmail.com
Phone: 8427112661, 9781810380

Corresponding author: Manorama Setia

ABSTRACT

Channel estimation is an important technique at receiver’s end in order to combat the effect of noise or interference on the transmitted information. The channel estimation reduces the effect of the channel on signal and recovers the transmitted signal. These estimation techniques depend on the propagation model through which it is travelling. Estimation techniques LS, MMSE, PERFECT performs differently in different scenarios. Techniques like PERFECT and MMSE are complex but are good in performance, whereas LS has less complexity, but its performance is less than MMSE. This paper makes a number of contributions, as well as, deeply discusses the comparison of estimation technique in different realistic channel and further we analyse the effect of variation in interpolation scheme, and transmission diversity of antennas in order to make performance better.

Keywords: ICI; ISI; eNodeB; OFDM; LS; MMSE; CFR

African Journal of Computing & ICT Reference Format:

I. INTRODUCTION

With the sudden boom in wireless applications, the demands for high data rates and high network availability are also touching the new peaks. OFDM is a promising technology to fulfill these demands due to its bandwidth efficiency and capability of high data rate transmission. OFDM is a technology which makes frequency selective channel into the flat fading channel. Channel estimation is a vital technique used at the receiver side in order to estimate actual transmitted signal which gets affected by ICI and ISI. Channel estimation technique is of two types i.e. blind and non-blind. The blind technique considers only statistical properties while in non-blind technique it considers some random part or even can consider whole part of the signal. [1] LTE system is MIMO-OFDM based and uses cyclic prefix. By using cyclic prefix at the beginning of the symbol, it overcomes from Inter-Symbol Interference (ISI) and Inter-Carrier Interference (ICI). This cyclic prefix can be longer or can be as same of channel B.W. but, it can be shorter in length due to some unforeseen effects which will lead to the introduction of ISI and ICI due to which the time estimation becomes strenuous [2]. The Channel estimation is done by using pilot signals which is placed along the data by using it we can increase the performance but at the cost of bandwidth efficiency.

The Pilot symbol should be transmitted in sufficient amount in order to make the receiver to estimate the transmitted information accurately. In the downlink only single pilot symbol can be used by terminals which are transmitted by eNodeB (evolved NodeB) and for uplink each terminal needs to send pilot separately. In order to make channel estimation as band efficient, the only option is to reduce the size of the pilot symbol. In channel estimation field, an adaptive method for polynomial interpolation based on Lagrange polynomial is launched during the conference held in 2014, which is helpful in making the lower order interpolation scheme like linear to outperform and retaining the computational complexity as low as possible [9].

Problems come under designing of channel estimators are how to arrange pilot information and how to make a suitable estimator for a system with less complexity [3]. A good channel estimation technique is important to recover signal by demodulating the received signal efficiently and accurately. As these techniques are behaving differently in different realistic channels environment so in this paper we are going to discuss how these estimation techniques vary in different realistic channels and within this estimation techniques how other parameters effect like changing interpolation techniques, transmit diversity how all these parameters are going to affect the performance.
In section II we will discuss about the background of parameters of which we are taking their effect. In section III we are discussing the environment of simulation, section IV consists of results obtained by simulations in the simulator and lastly we conclude the results.

2. BACKGROUND

In order to increase the data rate, MIMO-OFDM technology is introduced but it not only introduced for the sake of data rate but also because it is flexible and reliable technology and behaves as a robust technology for the multipath fading, which a biggest problem of channels. [3]

![Figure 1: channel estimation process][1]

Channel estimation technique is basically to overcome the loss of data occurred in the region free out of ISI, in which we can apply blind and non-blind techniques but in today’s scenario blind techniques are avoided and more reliable communication relay on data aided channel estimation technique. In these techniques estimation is accompanied with interpolation technique and from figure 1, process can be analysed.

2.1 Estimation techniques

2.1.1 Least square

The key role of this estimator is to reduce the distance between transmitted and received signal. LS is basically used as a initial stage for other complex estimator. LS channel estimator can be expressed as in eq. (1):

\[ H_{ls} = X^{-1}Y \] (1)

It ignores the effect of AWGN noise but is widely used because of its less complexity, but it is not a reliable technology to estimate the original signal from received one. [2] LS estimation is basically used at initial stage for the estimation of pilot subcarriers and after words it will be expanded for others subcarriers through interpolation [4]

2.1.2 Minimum mean square error

It is known by MMSE as well as LMMSE (Linear minimum mean square error) it is more good in performance as compared to least square because it considers the effect of AWGN noise introduced when signal is passing through channel. But it requires extra information for estimation like channel correlation and SNR [4] and this making up backup of channel information, makes this technique complex. It actually reduces the MSE which is as in eq (2):

\[ \text{MSE} = E \{ |He - H|^2 \} \] (2)

He = estimated attenuation vector
H = true with attenuation vector

2.1.3 Perfect channel estimation

In this estimation technique CFR (channel frequency response) is known by receiver in advance. [6]

2.2 Interpolation techniques

2.2.1 Linear interpolation

It is widely used technology for practical implantation because of its less complexity and good performance. It considers the variation between the two pilot sub-carriers which is expressed by CFR (channel frequency response) is uniform, because of this, output comes out to be linear [4]

2.2.2 Higher Order Interpolation Techniques

Linear interpolations come under category of first order interpolation they perform well but to improve the efficiency by using smoother fittings which can approximate the realistic channel for improving the efficiency, high order implementation come into existence but at the cost of complexity. It is good in accuracy but for slow variations in channel its accuracy is not improved. These interpolation techniques generally use more than two pilot subcarriers while some require more channel information. Some examples are spline, cubic, Gaussian etc. [4]

2.3 MIMO modes

As MIMO modes has different modes Transmit diversity and spatial multiplexing. It can be used both on the transmitter as well as on the receiver side. On the receiver side, we are only getting the replica of information and from that we have to recover the original signal. In the transmit diversity we transmit same data based on STBC codes. Here we are considering only transmit diversity modes in which we can take 2*1 or 4*2 modes as first no represent transmitting antennas and second no. represent receiving antennas through which error rate becomes low. This scheme increases the reliability by reducing the error rate. [5]

3. SIMULATION ENVIRONMENT & METHODOLOGY

The vital part of wireless communication is environment in which our data is travelling and in a practical consideration there is no clear path between transmitter and receiver, due to this interference signal Values of Parameters Used for Simulating Environment goes under path losses with reflection, refraction and scattering, and due to which signal follows different path to reach the receiver, so there is an average delay on the way, therefore ITU has standardised these to make real analysis of the simulation. [7]

3.1 Pedestrian A and B

Pedestrian environment is made up of small cells and low transmit power. In this path loss rule is considered which is \( R^{-\alpha} \) and if an interference free area (LOS) is their then we take \( R^{-\alpha} \) as path loss. Antenna with low height are located in outdoors and users in pedestrian environment are located on the streets and in covered areas mobile speed is assumed to be 3Km/hr. The Pedestrian A model has 4 and Pedestrian B model has 6 paths. For getting the relative and path loss we can refer to ITU guidelines. [8]
3.1.2. Vehicular A

Vehicular environment is made up of large cells and higher transmit power. Hence, it consumes more bandwidth and capacity becomes the area to focus. Vehicular speed is assumed to be 120Km/hr but due to path loss occurrence with increase in distance power of signal also decreases. As the path loss exponent is 4 for both urban and suburban areas which will be comparatively less for rural areas. [10]

3.2. Environment values

The parameters set for the simulation is the realistic environment made for analysing the results which further help in making practical implementations. These parameters include no. of sub-frames, Tx- diversity, simulation type, spacing between subcarriers, size of CP, HARQ, filtration used etc.

Table1: Values of Parameters Used for Simulating Environment

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of sub-frames</td>
<td>500 or 1000</td>
</tr>
<tr>
<td>CQI</td>
<td>7</td>
</tr>
<tr>
<td>Modulation type</td>
<td>16 QAM</td>
</tr>
<tr>
<td>Transmit diversity</td>
<td>111,221,242</td>
</tr>
<tr>
<td>HARQ</td>
<td>0</td>
</tr>
<tr>
<td>Channel estimation</td>
<td>LS, LMMSE, PERFECT</td>
</tr>
<tr>
<td>Channel interpolation</td>
<td>Linear, spline, cubic, T-F, sinc_freq</td>
</tr>
<tr>
<td>Spacing between subcarriers</td>
<td>15e3</td>
</tr>
</tbody>
</table>

3.3. Methodology

1. Set the global parameters debug level to highest level.
2. Select the path environment by selecting ITU standardised channel.
3. Decide the transmission whether it is MIMO or SISO transmission with the help of transmission mode variable.
4. Set the variable of channel estimation
5. Switch channel estimation method

Case LS
- Locate the reference symbols
- Convert position of reference symbol into appropriate matrix
- Then refer symbol in terms of frequency

Case MMSE
Firstly LS estimate on the positions of symbols, then again same procedure of reference symbols

Case PERFECT
In this all knowledge of channel is with estimator in advance.
- Channel interpolation is considered with channel estimation Linear, Spline, and Cubic.
- Sub-frames start transmitting within the range of SNR.
- The results will be traced in terms of BLER and throughput.

4. LINK LEVEL RESULTS AND DISCUSSION

In this section we are going to analyse the results after evaluating certain environment conditions through LTE link level simulator for different channel estimation techniques and making it better by changing some other parameters like transmit diversity, channel filtration, channel interpolation technique for different realistic channels (PedB, VehA). In fig. 1 and 2 we are analysing the BLER and throughput of different estimation technique from fig. 1 we can observe PERFECT estimation technique is performing better than LS and MMSE as up to SNR=0, all are showing 100% error. With increase in SNR, PERFECT perform good as it require 2-2.5 db less power as compare to LS and 1db w.r.t MMSE where as in VehA it takes 1db more up to 15db after that this technique perform good for VehA. Same for MMSE but for LS both channel perform same. In table 1 and 2 shows the BLER and throughput performance for different estimation technique.
Table 2: BLER Results of Simulation of Realistic Channels for Different Estimation Technique

<table>
<thead>
<tr>
<th>Channel estimation</th>
<th>BLER $10^{-2}$</th>
<th>PedB</th>
<th>VehA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perfect</td>
<td>11db</td>
<td>11.5db</td>
<td></td>
</tr>
<tr>
<td>MMSE</td>
<td>12 db</td>
<td>12.5db</td>
<td></td>
</tr>
<tr>
<td>LS</td>
<td>13 db</td>
<td>13.5db</td>
<td></td>
</tr>
<tr>
<td>Perfect</td>
<td>17db</td>
<td>16db</td>
<td></td>
</tr>
<tr>
<td>MMSE</td>
<td>16.5db</td>
<td>17db</td>
<td></td>
</tr>
<tr>
<td>LS</td>
<td>19db</td>
<td>18db</td>
<td></td>
</tr>
</tbody>
</table>

From Fig. 2, in this also PERFECT performs well, as compare to other estimation technique only for PERFECT PedB is requiring 1db less power than VehA for other both are performing same and in this LS is performing by consuming 2.5-3 db less power as compare to LS.

Table 3: Throughput Results of Simulation of Realistic Channels for Different Estimation Technique

<table>
<thead>
<tr>
<th>Channel estimation</th>
<th>Target Throughput (Mbps)</th>
<th>PedB</th>
<th>VehA</th>
</tr>
</thead>
<tbody>
<tr>
<td>PERFECT</td>
<td>1</td>
<td>8db</td>
<td>8.5db</td>
</tr>
<tr>
<td>MMSE</td>
<td>9db</td>
<td>8.75db</td>
<td></td>
</tr>
<tr>
<td>LS</td>
<td>10db</td>
<td>10db</td>
<td></td>
</tr>
<tr>
<td>PERFECT</td>
<td>1.2</td>
<td>10db</td>
<td>11db</td>
</tr>
<tr>
<td>MMSE</td>
<td>11.5db</td>
<td>11.35db</td>
<td></td>
</tr>
<tr>
<td>LS</td>
<td>13 db</td>
<td>13db</td>
<td></td>
</tr>
</tbody>
</table>

From Fig. 3 and 4, representing the variation come when we use filtration for channel either block fading or fast fading in this we consider the effect of filtering on poor performance estimation.

Figure 2. Throughput analysis for different estimation technique for VehA, PedB

Figure 3. BLER analysis for various interpolation techniques for PERFECT technique with PedB channel

In this we have taken estimation technique as PERFECT in which we compared different interpolation technique which proofs if higher order spline or T-F technique is used BLER performance becomes better by 5db which is a great improvement, this is also same for throughput.
Throughput analysis for various interpolation LS estimation with PedB channel

In Fig. 5 and 6 shows that for LS estimation variation in interpolation shows very less effect on BLER and throughput than cubic performs in terms of BLER up to 11 db max. 1db power saving is achieved. The same is the achievement for throughput.

Table 4: BLER Results of Simulation of Realistic Channels for Different Transmit diversity

<table>
<thead>
<tr>
<th>BLER</th>
<th>Transmit diversity</th>
<th>MMSE (SNR)</th>
<th>LS (SNR)</th>
<th>PERFECT (SNR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10^{-2}$</td>
<td>2*1</td>
<td>8.5db</td>
<td>9db</td>
<td>8db</td>
</tr>
<tr>
<td></td>
<td>4*2</td>
<td>4db</td>
<td>6db</td>
<td>3.5db</td>
</tr>
<tr>
<td>$10^{-2}$</td>
<td>2*1</td>
<td>11db</td>
<td>12db</td>
<td>10.5db</td>
</tr>
<tr>
<td></td>
<td>4*2</td>
<td>5.5db</td>
<td>-</td>
<td>4db</td>
</tr>
</tbody>
</table>
Figure 5. BLER analysis for various interpolation techniques for LS estimation with PedB channel

Figure 6. Throughput analysis for various interpolation techniques with PedB channel
In Fig. 7 and 8 we can analyse that as how different transmit diversity effect different estimation technique like LS, MMSE, PERFECT which shows with increase in transmit diversity for simple estimation technique like LS makes it better by 6-7db which is a good achievement in saving the power dissipation and in terms of throughput 3-4db power is saved.

Table 3 and 4 shows the BLER and throughput performance of different transmit diversity.

Figure 7. BLER analysis for different estimation technique for different transmit diversity in PedB

Figure 8. Throughput analysis for different estimation technique for different transmit diversity in PedB
Table 5: Throughput Results of Simulation of Realistic Channels for Different Transmit diversity

<table>
<thead>
<tr>
<th>Target Throughput (Mbps)</th>
<th>Transmit diversity</th>
<th>MMSE (SNR)</th>
<th>LS (SNR)</th>
<th>PERFECT (SNR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2*1</td>
<td>7db</td>
<td>8.5db</td>
<td>6.7db</td>
</tr>
<tr>
<td>1</td>
<td>4*2</td>
<td>3.5db</td>
<td>5.5db</td>
<td>2.5db</td>
</tr>
<tr>
<td>1.2</td>
<td>2*1</td>
<td>10db</td>
<td>11db</td>
<td>9db</td>
</tr>
<tr>
<td>1.2</td>
<td>4*2</td>
<td>5db</td>
<td>7db</td>
<td>5db</td>
</tr>
</tbody>
</table>

The same results are observed for VehA channel width minute difference of 0.1-0.5 db comparatively.

5. CONCLUSION

The channel estimation is an important technique to combat the effects of ISI and ICI which get introduced when signal is travelling through realistic channel. This is data aided estimation with standing interpolation. In this paper we concluded from our observations obtained, and realistic environment simulation results, that PERFECT is best technique among all estimation techniques, but it requires channel impulse response, which shows that it is a technique with more load but by using spline or T_F interpolation we can save 5-6db power just with first order estimation only. We have also observed that as compared to SISO, by using MIMO good performance can be achieved.

FUTURE SCOPE

There are a lot of possibilities in this area of research. This research can be further enhanced by researching evaluating estimation techniques for extended realistic channels (EVehA and EPedB) and the performance can further be enhanced by using CLSM as compared to transmit diversity with which performance of less complex techniques can also be improved to a much greater extent.

REFERENCES


Author’s Brief

Manorama Setia is currently pursuing M.tech in Electronics and Communication with specialisation in communication from Guru Nanak Dev University, Regional Campus- Jalandhar. She received her B.Tech degree from Rayat and Bahara College of engineering, Mohali in 2013. Her research areas of interests includes Wireless Communication.

Dr. Jyoteesh Malhotra received B.Eng., M.Tech, PhD, is involved in teaching and research at Electronics and Commun. Department, Guru Nanak Dev University, Regional Campus, Jalandhar. His research interests includes LTE, simulation of wireless channels, Fading mitigation techniques, and wireless networks. Dr. Malhotra has published more than 90 research papers in the International Journals/ Conferences and authored 02 books. He is in editorial board of many International Journals of repute.
A Mobile Architecture For Childhood Vaccine Preventable Illnesses Expert System

K. Obahiagbon
Department of Mathematics and Computer science
Benson Idahosa University
Edo State, Nigeria.
E-mail: kobahiagbon@biu.edu.ng
Tel: 08131130870

B.B. Odigie
Department of Management Information System
uchi Polytechnic, Edo State, Nigeria.
E-mail: ditsciences@yahoo.com
Tel: 08023314521

ABSTRACT

Globally, vaccination against childhood communicable diseases through the Expanded Program on Immunization (EPI) is one of the most cost-effective public health interventions put in place by various governments to eradicate and remedy the menace caused by to childhood illnesses and diseases. Despite government continuously efforts to provide these vaccines free, most pregnant and nursing mothers especially in developing countries are ill-informed of the gains and needs of participating in effect vaccination of the young child. Also, the limited number of qualified health care providers against the increasing population, poor access to quality health care facilities, poor sanitary culture and lack of electronic medical records of patients that have received any form of vaccinations had hampered the smooth delivery and administration of these vaccines in developing countries. To this end, searching for innovative technologies and methodologies for the administration and universal coverage of child immunization will among other processes improve the quality of services given by health care provider to both mother and child. This research blends child vaccination processes with the innovative power of telemedicine to provide a software system framework that will both provide an alert mechanism on available child vaccine for a specific period and advisory services to nursing mothers. It will help to furnish health care provider with medical decisions and data for effective vaccination planning. It will also assist the government to budget properly towards sustaining effective child vaccination based on the evidence available in the centralized Electronic Medical Records.

Keywords: Effective Immunization, Immunization coverage, Electronic Medical Record.

African Journal of Computing & ICT Reference Format:

I. INTRODUCTION

The global menace of vaccine preventable illnesses and diseases such as Diphtheria, Tetanus, Pertussis (whooping cough), Measles, Smallpox and Polio has accounted for over 29% of death of children under five (5) years. Effective Immunization was recorded as the most successful public health initiatives and it has helped to avert an estimated 2-3 million life-threatening diseases and death that disproportionately affect children [1]. UNICEF reported that 123 countries had immunized over 90% of infants against measles in 2011 and vaccination had resulted in a 71% drop in measles deaths worldwide between 2000 and 2011. UNICEF added that new promising vaccines for children against Rotavirus, Haemophilus influenza B, Hepatitis B and Rubella diseases had been introduced in 31, 177, 180, and 130 countries respectively.

Approximately 82% of newborns in these countries were protected against neonatal tetanus through immunization. Despite these amazing track successes recorded, many countries are yet to introduce important vaccines such as the rotavirus and pneumococcal vaccines because of insufficient funding, underperforming cold chains or difficult access to vulnerable children who resides in rural areas. More also, unavailability of vaccines, poor or inaccessible health services because of increasing population, poor information on immunization (when and why patient most bring their children for immunization) have resulted to over 30 million children unimmunized worldwide and over 70% of these unreached children lived in Afghanistan, Chad, the Democratic Republic of Congo, Ethiopia, India, Indonesia, Nigeria, Pakistan, Philippines and South Africa [2].
Dr Temi Adekunle highlighted challenges inhibiting the effective coverage of immunization as inadequate awareness and sensitzation especially in remote locations, poverty, limited funding, poor healthcare system and lack of adequate monitoring [4]. One strategy to properly inform families, increase awareness and provide updates of immunization coverage to the health care system globally is to create or design innovative methodologies and technologies that will facilitate the collection and dissemination of immunization information remotely and intelligently.

This can be achieved by enabling wireless and web based applications that collect vital health information from patients (pregnant mothers and newborn babies) irrespective of their location to a centralized Electronic Medical Record Server database or centre; and an expert system who intelligently analyzes the transmitted information and provide informed decisions and recommend appropriate plan of actions for patients. Harnessing these opportunities will accelerate the global immunization coverage and will improve the healthcare systems thereby building applications that will be integrated into existing medical processes in designated healthcare institutions that will be accessible by mobile users. It will also afford collaborations between the developing country health care system and the developed countries. In this research, we propose a software framework that will blend information on vaccine preventable illness administration, treatment and control with telemedicine powered by Information and Communication Technology (ICT) to provide an immunization alert and expert system to families globally.

2. LITERATURE

2.1 Background of study

Child mortality has fallen significantly in many low income countries but however, sub-Saharan Africa continues had experienced the slowest fall in mortality rate among children[5]. UNICEF and WHO in the state of the world’s children report noted that 8.1 million children across the world who died in 2009 before their fifth birthday lived in developing countries and died from a disease or a combination of diseases that could easily have been prevented or treated [3]. It also noted that, half of these deaths occurred in just main five countries- India, Nigeria, the democratic republic of Congo, Pakistan and China- with India and Nigeria both accounting for one third of the total number of under five deaths worldwide. The report described the declining rate as disturbing and grossly insufficient to achieve the MDG goal by 2015 as only 9 out of the 64 countries with high child mortality rate are on track to meet the MDG goal [6]. Effective communication among communities was recommended by UNICEF as a measure that will help curb childhood illnesses and deaths [1].

[7] anticipated that scaling up the use of existing vaccines in 72 of the world’s poorest countries could save 6.4 million lives and avert $6.2 billion in treatment costs and $145 billion in productivity losses between 2011 and 2020. Models of vaccine-preventable disease burden frequently include immunization coverage levels among their components [8, 9,10].

Immunization coverage levels and trends are therefore indicators for monitoring the performance of immunization services locally, nationally and internationally and it provides guided strategies for eradicating, eliminating and controlling vaccine-preventable diseases. It also identifies areas of immunization systems that may require additional resources and attention. It assesses the need to introduce new vaccines into national and local immunization systems [11]. Effective coverage levels of vaccine preventable diseases and deaths are global indicators for tracking progress towards achieving Millennium Development Goal 4-to reduce child mortality and it provides a viable blueprint for immunization service delivery and disease control [31].

Currently, Nigeria is among the ten countries in the world with vaccine coverage rates below 50 percent and half of the 5% of death of children caused by measles in Africa occurs in Nigeria [5,12]. The country also has the highest prevalence of circulating wild poliovirus in the world [13,14].

2.2 Challenges of immunization coverage in developing countries

The developed countries already have their experiences of effective immunization coverage [11]. In developing countries, [15] stressed that healthcare problems resulting from different factors such as economic; poor planning or poor implementation of health policies; problem of availability, accessibility, affordability, and sustainability of health facilities; and weak referral system had plagued e-health system. He suggested that the use of communication equipment like mobile phones will help facilitate referral system. However, low standard of living and poor sanitary practices have increased child mortality caused by vaccine preventable blight and death in some developing countries [6].

As a panacea, [16] proposed an intervention package component of Performance Management and Demand Creation using SMS reminders to patients and staffs. Also, embracing innovations in telemedicine and e-health services will provide low cost access to health care facilities and effective instant SMS alerts messages and health tips on proper health styles to patients irrespective of their locations from the health care providers will help improve health care delivery [17].
2.3 Vaccine-preventable childhood diseases
According to [1], the common vaccine preventable childhood diseases were:

**Diphtheria:** a serious disease caused by a poison made by bacteria. It causes a thick coating in the back of the nose or throat that makes it hard to breathe or swallow.

**Hepatitis B:** a serious infection that affects the liver. Hepatitis B causes chronic liver disease and put people at high risk of death from cirrhosis of the liver and liver cancer.

**Haemophilus influenzae type B (Hib):** Hib can cause severe pneumonia, meningitis and other serious diseases almost exclusively in children under the age of 5.

**Measles:** a highly contagious respiratory disease caused by a virus. Measles causes fever, runny nose, cough and rashes all over the body. About one in 20 children with measles also gets pneumonia.

**Pertussis (whooping cough):** a highly contagious respiratory disease, which produces violent, uncontrollable coughing which often makes it hard to breathe. Pertussis most commonly affects infants and young children and can be fatal, especially in babies less than 1 year of age.

**Pneumococcal:** a disease that caused pneumonia, meningitis, or blood infection. Polio (poliomyelitis) mainly affects children under five years old. One in 200 infections leads to irreversible paralysis.

**Rotavirus** is the leading cause of severe diarrhea in infants and young children. Globally, it causes more than half a million deaths each year in children under 5.

**Congenital Rubella syndrome (CRS)** is a lifelong disabilities disease. Children whose mothers have rubella during the early stages of pregnancy often contract (CRS) and are at risk for other developmental problems such as congenital heart disease and mental retardation.

**Tetanus** is an extremely deadly and paralyzing disease. Mothers and newborns contract tetanus when deliveries happen in unhygienic conditions – as can be the case in remote and underdeveloped area Tetanus is also a fatal disease. It is a bacteria causing weakness and paralysis when allows to fester in a deep dirty wound.

**Tuberculosis (TB)** is a disease that typically attacks the lungs. Tuberculosis causes pulmonary infection but can spread to many other organs causing serious illness, death and disability.

**Yellow fever** is found in tropical climates and is transmitted to humans by the bite of an infected mosquito. Up to 50% of people who develop severe illness and are not treated may die.

2.4 Vaccination Schedule
[4] reported that the Bacilli Calmette Guerin (BCG) is administered to a child at birth or as soon as possible after birth. Oral Polio Vaccine (OPV) is given at birth and also at six, 10, and 14 weeks of age. OPV can also be received any other time, especially during the immunization weeks. Pentavalent vaccine is a combination of five vaccines-in-one that prevents diphtheria, tetanus, whooping cough, hepatitis b and haemophilus influenza type b, all through a single dose. She added that the Hepatitis B vaccine is given at birth, six and 14 weeks; Measles vaccine is administered at nine months of age and Yellow Fever, also at nine months. The Pneumococcal Conjugate Vaccine (PCV10) is a vaccine that prevents diseases caused by the pneumococcal bacteria.

OPV10, Chicken pox (optional) and Measles vaccine (MMR) are administered between 15 to 18 months. MMR is the measles, mumps and rubella vaccine. Measles, mumps and rubella are very common highly infectious, conditions that can have serious, potentially fatal complications, including meningitis, swelling of the brain (encephalitis) and deafness. Meningitis and Typhoid fever (optional) are administered after 24 months. Meningococcal vaccine is a vaccine used against Neisseria meningitis, a bacterium that causes meningitis, meningococcal. Septicemia and rarely carditis [18].

2.5 Related Literature
According to [19], the modern healthcare system focused more on patient centred approach than traditional hospital based system. The adoption of Information and Communication Technology (ICT) in health sector have powered amazing health innovations such as telemonitoring and remote monitoring and these innovations have demonstrated a 50% reduction in mortality and over 50% in productivities within the health care systems [20,21].

Telemedicine consists of the transmission of vital health data from a remote location to another location for data interpretation and decision-making and it has increased the quality of life in heart failure, diabetes mellitus, rhythm disorders, psychiatric diseases and blood pressure [22,23,24]. Remote monitoring is an interesting and useful area that needs a lot of research in order to come up with solutions that benefits humanity- increase the monitoring performances while looking for economical solution and reduce inconveniences to patients [25].

A remote medical monitoring system using sensor and GPRS to collect and process physiological data from patients with a view to providing medical advice and decisions was proposed by [26]. [27] developed an ambient-intelligent patient room where the nurse interacts intelligently with the patients based on evidence from patient’s surroundings/environments. A paper titled “Mobile Application for Diabetes Control in Qatar” was presented by [28] and it reflected on using mobile technology and applications to help diabetes patient in Qatar to manage their disease using glucose monitoring and diet management. [29] designed a secure based mobile healthcare system where patient data are encrypted using elliptic curve...
algorithm which will be decrypted in the mobile phone. A robust framework for intelligent remote blood pressure monitoring and control was proposed by [17] such that patient vital signs are collected remotely, analyzed by an expert system and intelligently classifying the patient blood pressure.

3. SYSTEM ARCHITECTURE

Base on evidences that effective immunization communication among communities and health care providers can help address the challenges of vaccine preventable childhood illnesses, a call for a systematic approach using best software modeling methodology became very paramount.

The proposed system consists of four modules as depicted in figure 1; the patient/user model, the doctor module, the health institution module and supporting partners module.

- The patient module (see figure 2) includes the devices for collecting physiological data from patients and a Remote User Interface (RUI) application where the patients can access the server if they have Internet service. The devices include medical related devices like: glucometer for measuring the blood glucose level, thermometer for measuring temperature, and weight scale; a java enable mobile phone to lunch the RUI application and connect to the server where they can instantaneously receive immunization health tips and SMS alert based on their health status.

Figure 1: Proposed system architecture
Figure 2: Overview of Patient Module

Figure 3: Overview of Doctor Module

Figure 4: Overview of Hospital Module
Doctor module includes all health care professionals, nurses and other health workers that are involved both in immunization and performing of clinical services as shown in figure 3.

Health Institutions module consist of Hospitals and the Electronic Medical Record (EMR) server (see figure 4). The EMR server has the expert system for effective communication and decision on immunization issues. The EMR was maintained and analyzed by an expert system and it controls the entire system. The EMR was implemented using open source database-MySQL- and JAVA programming language. The knowledge base of the EMR SERVER is where the physiological parameter will be analysed in the form of rules; each rule represents particular childhood disease knowledge which is based on the age of the patient. The inference engine looks for rules with conditions that will result to true based on the given fact (data from the user) and fire such rule.

The supporting partners module are organization and international partners actively involved in immunization such as WHO, UNICEF, Immunization Partners in Asia Pacific (IPAP) (see figure 5).

4. SYSTEM MODELING
According to [30], the Unified Modeling Language (UML) is the de-facto for object oriented design methodology and it depict the functionality of a system. UML was used to capture and model most of the functionalities in the proposed system. Two main UML diagrams- Structure diagrams and behavior diagrams was applied. Structure diagrams include class diagrams, object diagrams, component diagrams, and deployment diagrams and they are used to describe the relationship between classes. Behavioural diagrams describe the interaction between the users and the use case. Bahavioural diagram includes use case diagrams, state-chart diagrams, collaboration diagrams, sequence diagrams and activity diagrams. The Use case captures the standard flow of events within the system (see figure 6). In the diagram, there are two types of users: the remote patient who is wearing a wireless sensor and the healthcare experts in an ICT healthcare institution. Restriction using password was used to limit access to information on the EMR server.

Automation of Short messaging service (SMS) alerts sent on immunization vaccines and special health tips are sent to patient mobile devices from the medical experts system. The system’s operations and interactions among subunits are depicted using Sequence Diagram (see figure 7). The user logs in with unique ID generated by the EMR. This will enable the user to be connected to the system which is equipped with WAP application that runs on the devices and converts it to readable format that will be transmitted via gateways. The application establishes an internet connection to the database web server in the EMR.
5. SYSTEM IMPLEMENTATION AND DEPLOYMENTS

This software system consists of Remote Users Interface (RUI) application and Web Enabled Vaccine Preventable Expert System (WEVPES) application. Both applications were implemented using JAVA Software Development Kit (SDK) 7.0 and Netbeans 8.0 was the Integrated Data Environment (IDE) used for the development of the server and Remote User Interface (RUI). mySQL was the database server used for holding the EMR. The WEVPES was deployed and tested on HP Compag 500B computer system with windows 7 operating system and XAMPP Version: 1.8.3 server installed and configured for proper functionality.

The Remote User Interface (RUI) is a java archive file (jar) that runs in various java enabled mobile devices. For the purpose of this work, a mobile phone simulator was first used to execute the jar file and all patient signs from the simulator were sent to the hospital EMR database. Later a Nokia 5300, Nokia 502, TECNO phones was used to run the RUI and connections were established to the server via internet/Bluetooth connection. The minimum hardware configuration for WEVPES to run is a Pentium IV system with 512MB RAM and 250GH hard disk running windows XP or later version. The minimum hardware configuration for RUI to run is any portable mobile java enabled device with appropriate Bluetooth/internet connections.
6. SYSTEM TESTING AND REPORT GENERATION

6.1 Web Enabled Vaccine Preventable Expert System (WEVPES) Interface

The screenshots for WEVPES application are depicted in figure 8a-d

![Figure 8a: WEVPE home Page](image1)

![Figure 8b: WEVPE Book An appointment](image2)
Figure 8c: WEVPE mobile data queue

Figure 8d: WEVPE Geo-location of patient based on address from GUI
6.2 Remote User Interface (RUI) Interface
The screenshots for RUI application are depicted in figure 9a-e.
7.0 CONCLUSION

In this study we demonstrated the gains of improving effective immunization coverage in developing country like Nigeria through mobile remote monitoring. It was glaring that remote monitoring will help improve better health care delivery towards meeting the yearning of citizens who cannot easily access health institutions because of distance, and limited health care providers. For critical situation, the patient’s location can be geo-located using the address supplied by the users. Also, special health SMS alert are instantly sent by the EMR based on the age of user supplied during remote monitoring from the RUI. The system is an interactive expert system that provides a more precise solution that will address immunization coverage problems, maintained a centralized database that track the progress of immunization schedule and inform physicians of current health condition of patients. The system can be used as a tool to help ease the work of physicians and combat childhood preventable illnesses and death.

REFERENCES


Fish Classification Using Support Vector Machine

S.O. Ogunlana, O. Olabode, S.A. A. Oluwadare & G. B. Iwasokun
Department of Computer Science
Federal University of Technology
Akue, Nigeria

ABSTRACT

Fish recognition is presently a very complex and difficult task despite its commercial and agricultural usefulness. Some of the challenges facing accurate and reliable fish recognition include distortion, noise, segmentation error, overlap and occlusion. Several techniques, which include K-Nearest Neighbour (KNN), K-mean Clustering and Neural Network, have been widely used to resolve these challenges. Each of these approaches has inherent limitations, which limit classification accuracy. In this paper, a Support Vector Machine (SVM)-based technique for the elimination of the limitations of some existing techniques and improved classification of fish species is proposed. The technique is based on the shape features of fish that was divided into two subsets with the first comprising 76 fish as training set while the second comprises of 74 fish as testing set. The body and the five fin lengths; namely anal, caudal, dorsal, pelvic and pectoral were extracted in centimeter (cm). Results based on the new technique show a classification accuracy of 78.59%, which is significantly higher than what obtained for ANN, KNN and K-mean clustering-based algorithms.

Keywords: Support Vector Machine, Image Recognition/Classification

I. INTRODUCTION

Fish recognition is the act of recognizing or identifying fish species based on their features. It is also a process of identifying fish targets to species based on similarity to images of representative specimens [1]. Fish recognition is necessary for a number of reasons, which include pattern and contour matching, feature extraction, determination of physical or behavioral trait and statistical and quality control of fish species [2]. Fish recognition is also beneficial to fish counting and population assessments, description of fish associations and monitoring ecosystems [3]. Accurate recognition of fish species is important, as there are often legal restrictions on fishing practices when their existence is considered threatened or endangered.

ML method consists of a range of approaches that rely on Artificial Neural Network (ANN), Fuzzy logic, K-Nearest Neighbor (KNN), K-means clustering and Support Vector Machine (SVM) [7]. SVM (also referred to as Maximum Margin Classifiers (MMC)) consists of a group of learning algorithms, originally developed by Vapnik [8]. It performs simultaneous minimization of the empirical classification error and maximization of the geometric margin. A SVM performs classification by constructing an N-dimensional hyper-plane that optimally separates data into two categories based on an algorithm that finds the maximum-margin hyper plane with the greatest separation between classes.

The instance at the minimum distances from the maximum-margin hyper plane constitutes the support vectors. In this paper, a SVM-based platform for fish recognition and classification is presented. Section II presents a review of relevant literature on fish classification and SVM, Section III focuses on the summary of some existing fish classification techniques while Section IV presents the design of the proposed system. Sections V and VI focus on the experimental study and discussion respectively.
2. RELATED LITERATURE

The authors in [9] presented a computer vision-based system that reliably classifies different fish species based on length measurement and weight determination. The system has capability for using vision-based catch registration for automatic classification of fish species but requires great computer power and very expensive computation. Furthermore, it is only applicable to fish length and weight measurements. In [10], an artificial neural network-based platform for fish species identification is presented. The platform uses several statistical methods such as discriminate function analysis and principal component analysis. Its limitation is its high false identification rate due to its reliance on feature extraction, hierarchical classification and tree construction as its core function. Although, the work achieved better accuracy compared to some other techniques, it is time consuming, computationally bulky and used one to one classifier and imbalanced datasets which are not sustainable for large number datasets. Lee et al. in [17] carried out shape analysis of fish and developed an algorithm for removing edge noise and redundant data point base on nine species with similar shape features. Decision tree was presented as a suitable method for high accuracy while the number of shape characters needed and how to use them depend on the number of species and the kind of species required. Experiments conducted on a given number of fish images of various species recorded very significant classification success.

Larsen et al. in [18] presented a shape and texture based fish classification method with several images and species. Shape and texture features were separated using active appearance model that is based on principal component scores and linear discriminate analysis. Rova et al. in [19] applied SVM algorithm to fish recognition and constructed a texture-based mechanism that distinguishes between the Striped Trumpeter and the Western Butterfish species. Two templates (one per specie) were built and each query image was warped to both templates for the texture-based classifier.

3. EXISTING FISH CLASSIFICATION ALGORITHMS

Fish classification is necessary for identification, marketability, pricing, consumption, scientific research and so on and the summary of some of the existing fish classification algorithms are presented below:

**Artificial Neural Networks (ANNs) Algorithm**

ANNs comprised of simple neurons with three basic elements: a set of synaptic weight, integration and activation function. The mathematical model of the neuron K is expressed as follows [1, 10, 12]:

\[ U_k = \sum_{j=1}^{N} w_{kj} x_j + b_k \]  
\[ y_k = \varphi(U_k) \]  
\[ x_j \] is the input signal, \( w_{kj} \) is the weight from the jth to kth neuron, \( b_k \) is the bias of the kth neuron, \( \varphi(.) \) is the activation function, and \( y_k \) is the output of the kth neuron. Several types of activation functions are used in ANN, but the sigmoid function is used as follows:

\[ \varphi(u) = \frac{1}{1+e^{-u}} \]

The sigmoid function generates a continuous valued output between 0 and 1 as the neuron’s net input goes from negative to positive infinity. Training the network involves the back-propagation algorithm and the goal is to find a set of connection weights that minimizes an error function. The back-propagation algorithm consists of following four steps:

- Compute the error derivative (EA) as follows:

\[ E_{Ax_j} = \frac{\partial E}{\partial x_j} = y_j - d_j \]  

E represents Error, \( y_j \) is the activity level of the jth unit and \( d_j \) is the desired output of the jth unit.

- Computes how fast the error change as the total input received by an output is changed as follows:

\[ E_{Ix_j} = \frac{\partial E}{\partial x_j} = \frac{\partial E}{\partial y_j} \times \frac{\partial y_j}{\partial x_j} = E_{Ax_j} y_j (1 - y_j) \]
- Computes how fast the error changes as the weight on the connection into an output unit is changed as follows:

$$E_{Wij} = \frac{\partial E}{\partial W_{ij}} = \frac{\partial E}{\partial x_j} \cdot \frac{\partial x_j}{\partial W_{ij}} = E_i x_j$$  \hspace{1cm} (6)

- Computes how fast the error changes as the activity of a unit in the previous layer is changed as follows:

$$E_{A_i} = \frac{\partial E}{\partial A_i} = \frac{\partial E}{\partial x_j} \cdot \frac{\partial x_j}{\partial A_i} = \sum_j E_{Wij}$$  \hspace{1cm} (7)

**Principal Component Analysis (PCA) Algorithm**

PCA is used to reduce the dimensionality of a dataset. Given that \( x \) is an \( n \times p \) observation mean-centered data matrix with \( n \)–observation of \( p \)-variables, the mean centering is expressed as \( x_i \rightarrow \bar{x}_i \). \( x_i \) is the \( i \)th element of the vector \( x \) and \( \bar{x} \) is the mean of its elements. The covariance matrix \( S \) of \( X \) is defined as [20]:

$$S = COV(X) \equiv \frac{1}{n-1}XX^T$$  \hspace{1cm} (3)

\( N \) is the number of observation, \( X \) is the matrix and \( X^T \) is transpose matrix of \( X \). The first linear function is defined as \( Z_1 = \alpha_1^T x \). In PCA, the variance of the linear function \( Z_1 \) is maximize as follows:

$$Z_1 = \alpha_1^T x = \alpha_{11} x_1 + \alpha_{12} x_2 + \ldots + \alpha_{1p} x_p = \sum_{j=1}^{p} \alpha_{1j} x_j$$  \hspace{1cm} (9)

\( \alpha \) is the eigenvector of covariance.

The variance is maximize using the Lagrange multiplier \( \lambda \) as follows:

$$\frac{\partial}{\partial x} [f(x) - yg(x) - 1] = 0$$  \hspace{1cm} (10)

$$f(x) = \alpha_1^T x$$  \hspace{1cm} (11)

$$g(x) = \lambda$$  \hspace{1cm} (12)

The differential gives the eigenvalue for the covariance matrix as follows:

$$\alpha_{11} - \gamma I_p = 0$$  \hspace{1cm} (13)

\( I_p \) is a \( p \times p \) identity matrix and \( \gamma \) is the variance. The percentage variation by the corresponding principal component is calculated from the respective eigenvalues as follows:

$$P_{100} = \frac{\gamma_1}{\gamma_1 + \ldots + \gamma_p} \times 100\%$$  \hspace{1cm} (14)

**K-Nearest Neighbour (KNN) Algorithm**

Given a training set \((x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\) where \( x_i \) is \( d \)-dimensional feature vector of real numbers, for all \( i \), \( y_i \) is the class label for all \( I \), then the task is to find \( y_{\text{new}} \) from \( x_{\text{new}} \). KNN algorithm involves finding \( k \) closest training points to \( x_{\text{new}} \) with respect to the Euclidean distance. The Euclidean distance is defined as [21, 22]:

$$E = \sqrt{\sum_{i=1}^{k} (x_i - x_{\text{new}})^2} \hspace{1cm} (15)$$

\( a_1, a_2, \ldots, a_d \) represent the scaling factor for different dimensions. During testing, for new test data \( x_{\text{new}} = (x_{1\text{new}}, x_{2\text{new}}, \ldots, x_{d\text{new}}) \) the classifier generates label \( y_{\text{new}} = (y_{1\text{new}}, y_{2\text{new}}, \ldots, y_{d\text{new}}) \). The percentage of accuracy, \( A \) is obtain as follows:

$$A = \frac{a}{b} \times 100$$  \hspace{1cm} (16)

\( a \) is number of correctly classified items and \( b \) is the number of items respectively.

**K-means Clustering Algorithm**

K-means clustering is an algorithm that classifies or groups object into \( k \) number of group based on attributes or features. The grouping is based on minimization of the sum of squares of the distances between data and the corresponding cluster centroid.

The k-means clustering algorithm is presented as follows [23, 24]:

- Initialize cluster centroid \( \mu_1, \ldots, \mu_k \in R^p \) randomly.
- Repeat until convergence:

\( k \) is the number of cluster, \( \mu_i \) is cluster centroids.

The initialization is by randomly picking \( k \) training set while the inner-loop repeatedly assign each training set \( x_i \) to the closest cluster centroid \( \mu_i \) and move each cluster centroid \( \mu_i \) to the mean of the points assigned to it. The centroid coordinate is determined thus:

Given \( x^i \), the \( i \)th coordinate of \( x^{k+1} \) is given by \( x_i^{k+1} = \text{argmin} f(x_{1}^{k+1}, \ldots, x_{i-1}^{k+1}, y, x_{i+1}^{k+1}, \ldots, x_{p}^{k+1}) \); for \( y \in R \). Thus, an initial guess \( x_i^{0} \) is used for a local minimum of \( F \), and a sequence \( x_i^{0}, x_i^{1}, x_i^{2}, \ldots \) is obtained iteratively.
Using line search in each iteration, then \( F(x^k) \geq F(x^{k+1}) \) while the K-means convergence is achieved by defining the objective function:

\[
J(c, \mu) = \sum ||x_i - \mu_c||^2 \quad (19)
\]

The inner-loop of k-means repeatedly minimizes \( J \) with respect to \( c \) while holding \( \mu \) fixed and minimizes \( J \) with respect to \( \mu \) while holding \( c \) fixed and thus, the value of \( J \) converges.

4. PROPOSED FISH RECOGNITION SYSTEM

The architecture of the proposed system is presented in Figure 1. The system uses the support vector machine (SVM) process (shown in Figure 2) to analyze and then classify fish species according to their features or characteristics by constructing an N-dimensional hyper-plane that optimally separates fish species into categories. The hyper-plane is based on a predictor variable and a vector of predictor values (which is the set of values assigned to the different fields in the dataset) is formed. Most importantly, SVM modeling is used to find the optimal hyper-plane that separates clusters of vectors on a hyper-plane that separates clusters of vectors in such a way that cases with one category of the target variable (classified as the training subset) are on one side of the plane and cases with the other category (classified as testing set) are on the other side [4]. For this research, the training set is a collection of the features of the fish and it is used to discover the predictive relationship among the fish species while the testing set consists of the features actually extracted from the fish species and is used to access the strength of the SVM algorithm for correct classification. The SVM algorithm comprises of the following components:

![Figure 1: Proposed Fish Recognition Model](image)

Optimal Separating Hyper-plane (OSH)

With OSH, there is assumption that a set \( S \) of points \( X_1, X_2, \ldots, X_N \) exists such that each point \( X_i \) belongs to one of the two classes and thus is given a label \( y_i \in \{-1, 1\} \). The goal is to establish the equation of a hyper-plane that divides \( S \) leaving all the points of the same class on the same side while maximizing the distance between the two classes and the hyper-plane. The set \( S \) is linearly separable if there exist \( W \in \mathbb{R}^n \) for dataset of size \( T \) and \( b \in \mathbb{R} \) such that the separating hyper-plane is defined as follows [25]:

\[
y_i(W^T X_i + b) \geq 1 \quad \text{for} \quad i = 1, 2, \ldots, N, \quad (20)
\]

the pair \((w, b)\) defined a hyper-plane of the equation:

\[
W^T X_i + b = 0
\]

If the norm of \( w \) is denoted with \( w \), then the signed distance \( d_n \) of a point \( x_n \) from the separating hyper-plane \((w, b)\) is defined by:

\[
d_n = \frac{W^T x_n + b}{||w||} \quad (22)
\]

The integration of Equations 1 and 3 for all \( X \in S \) gives:

\[
\frac{1}{w} \quad (23)
\]

\( \frac{1}{w} \) is the lower bound on the distance between point \( x_n \) and the separating hyper-plane \((w, b)\). Given a separating hyper-plane \((w, b)\) for the linearly separable set \( S \), the canonical representation of the separating hyper-plane is obtained by rescaling the pair \((w, b)\) into the pair \((\frac{w}{||w||}, \frac{b}{||w||})\) in such a way that the distance of the closest point, say \( x_p \), equals \( \frac{1}{w} \) [25]:

\[
\min_{w,b} \{\min_{x \in S} \|W^T x + b\| = 1 \quad (24)
\]
A separating hyper-plane is the canonical representation and given a linearly separable set S, the optimal separating hyper-plane is the one with the closest distance (which in most cases is equal to $\frac{1}{w}$) to S.

**The Margin of Separation (MS)**

The margin of separation is the lower bound of the minimum distance between points of different classes. MS can be coupled with OSH to form a separating hyper-plane that maximizes the margin which can be thought of as a measure of the difficulty of the problem. In general, the quadratic optimization problem with linear constraints is expressed based on MS as:

$$\text{Minimize } \frac{1}{2} \|w\|^2 = \frac{1}{2} w^T w$$

Subject to $y_n (w^T x_n + b) \geq 1, n = 1, 2, 3, ..., N$ (25)

with respect to $w \in R^d$ and $b \in R$.

w is the Euclidean space R and b is a scalar belonging to real number.

The optimization problem has a convex quadratic objective and only linear constraints. Its solution, obtained by using the Lagrange Multipliers Technique (LMT), gives the optimal margin classifier as well as the separating hyper-plane with the best possible margin. The LMT for a multiplier $\alpha_n$ is defined as follows:

$$\text{Minimize } L(w, b, \alpha_n) = \frac{1}{2} w^T w - \sum_{n=1}^{N} \alpha_n [y_n (w^T x_n + b) - 1]$$

subject to $\sum_{n=1}^{N} \alpha_n y_n = 1, n = 1, 2, 3, ..., N$ (26)

$\alpha_n$ is a vector of non-negative Lagrange Multipliers.

The solution is obtained by solving the standard quadratic programming:

$$w = \sum_{n=1}^{N} \alpha_n y_n x_n$$

(27)

$$b = \sum_{n=1}^{N} (w^T x_n - y_n)$$

(28)

$$\sum_{n=1}^{N} \alpha_n y_n = 0$$

(29)

$N$ is the number of support vectors and only a few $\alpha_n$ is greater than 0, corresponding to the support vectors. The solutions to $w$, $b$, and $\alpha_n$ are still unknown. To solve for $\alpha_n$ $w = \sum_{n=1}^{N} \alpha_n y_n x_n$ and $\sum_{n=1}^{N} \alpha_n y_n = 0$ are substituted into $L(w, b, \alpha_n)$ to get:

$$L(\alpha_n) = \sum_{n=1}^{N} \alpha_n - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_n \alpha_m y_n y_m x_n^T x_m$$

(30)

Equation 7 becomes:

$$L(\alpha_n) = \sum_{n=1}^{N} \alpha_n - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_n \alpha_m y_n y_m k(x_n, x_m)$$

(31)

$k(x_n, x_m) = x_n^T x_m$ is a Kernel function.

Equation 12 is solved by quadratic programming to get global optimal of $\alpha_n$ by maximizing $L(\alpha_n)$ as follows:

$$\text{Maximize } L(\alpha_n) = \sum_{n=1}^{N} \alpha_n - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_n \alpha_m y_n y_m k(x_n, x_m)$$

subject to $\alpha_n \geq 0$ for $n = 1, 2, ..., N$ and $\sum_{n=1}^{N} \alpha_n y_n = 0$ (32)

5. EXPERIMENTAL STUDY

The experimental study of the proposed algorithm took place on a Brian System with Dual Core T5900 at 2.20 GHz processor, 2GB RAM and Window Vista 32-bit Operating System. MATLAB 2000b and Microsoft Access Database Management System featured as the frontend and backend engines respectively. The study was based on shape feature and image texture datasets obtained from the selected species through collaboration between Fishery Departments of the Federal University of Technology, Akure (FUTA), Nigeria and Adekunle Ajayi University, Akungba-Akoko (AAUA), Nigeria. Six features: namely body length, anal fin length, caudal fin length, dorsal fin length, pelvic fin length and pectoral fin length (see Figure 3) were extracted. The fish texture dataset comprises of extracted texture from the two species.

Each fish image was in JPEG format of 20x20 pixels at 256 grey levels per pixel. Fish classification involves separating data into training and testing sets with each instance of the training set containing one target value (class label) and several attributes (the features). Based on the SVM, a model that is based on the training data is produced for predicting the target values of the test data when only the test data attributes is given. The testing set was used to assess the strength of the SVM classification model. The sequence of operations on the data required by SVM for classification includes data preparation (which ensured that two classes of data were collected), data conversion to SVM format and determination of the class of the input feature based on the algorithm presented in Figure 4.
The summary of the parameters and the results for different algorithms with equal threshold is presented in Table 1. The use of these features is premised on their commonness and measurability. The system was trained using 76 fish consisting 38 Ethmalosa fimbriata and 38 Scomberomorus tritor. The Class Ethmalosa fimbriata was assigned 1 while the class Scomberomorus tritor was assigned 2. Seventy-four (74) fish with 37 Ethmalosa fimbriata and Scomberomorus tritor species apiece were used for testing. SVM method was based on a classifier with linear kernel due to the linearly separable nature of the data and a recognition accuracy of 78.59% was recorded. For Classification using neural network algorithm, the network was trained using back-propagation algorithm and a recognition accuracy of 60.01% was returned. The K-Nearest Neighbour (K-NN) classification algorithm used k=7 and a recognition accuracy of 52.69% was returned. The choice of k depends on the data and it must be odd number to avoid ties. Smaller k resulted in higher variance (less stable), while larger k resulted in higher bias (less precise).

Finally, the recognition accuracy for Classification using K-Means Clustering Algorithm was 50.97%. These results indicate superior performance for SVM algorithm in the classification of fish species.

Table 1: Experimental results for Fish Classification Based on different methods

<table>
<thead>
<tr>
<th>S/N</th>
<th>Parameters</th>
<th>Experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>SVM</td>
</tr>
<tr>
<td>1</td>
<td>No. of Fish/Observation</td>
<td>150</td>
</tr>
<tr>
<td>2</td>
<td>No. of training Set</td>
<td>76</td>
</tr>
<tr>
<td>3</td>
<td>No. of testing Set</td>
<td>74</td>
</tr>
<tr>
<td>4</td>
<td>No. of Fish Family</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>No. of input Features</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>Recognition Accuracy</td>
<td>74.32%</td>
</tr>
</tbody>
</table>
6. CONCLUSION

This paper contributed to knowledge by formulating an SVM-based fish classification algorithm. Six shape features; namely body length, anal fin length, caudal fin length, dorsal fin length, pelvic fin length and pectoral fin length were extracted from 150 fish (divided into 76 training and 74 testing sets) and the extraction formed the basis for the classification. The classification results exhibited the potential of the new algorithm for reliable and adequate fish classification and placed it at comparative advantage over some existing techniques such as ANN, K-NN and K-Means Clustering. However, the obtained recognition accuracy of 78.59% reveals there is still a lot of room for improvement. Future research therefore aims at improving the classification rate and performing fish classification on larger datasets and species.

![Flowchart of SVM Process](image.png)

Figure 4: Flowchart of SVM Process

REFERENCES


[14] Rova A., Mori G. and Dill L.M. (2012). One Fish, Two Fish, Butterfish, Trumpeter: Recognizing Fish in Underwater Video


[22] Padraig Cunningham and Sarah Jane Delany (2007), k-Nearest Neighbour Classifiers, Technical Report UCD-CSI


Performance Evaluation of Routing Protocols in MANETs using Varying Number of Nodes and Different Metrics

O.O. Omitola
Department of Electrical/Electronic and Computer Engineering
Afe Babalola University
Ado Ekiti, Nigeria
Email: omitolasegun@yahoo.com
Phone number: +2348077176985

ABSTRACT

This work evaluates the performance of three well known routing protocols in mobile ad-hoc networks, namely Ad-hoc On demand Distance Vector Protocol (AODV), Dynamic Source Routing (DSR) and Temporally Ordered Routing Algorithm (TORA). Two scenarios were created using OPNET and extensive simulations were made on each scenario to evaluate the performance of the three protocols. Results were presented as a function of three performance metrics namely: throughput, delay and network load. In most simulations and overall, DSR and AODV performed better than TORA while DSR shows a better performance compare to AODV. In terms of throughput, DSR is better than any of the other two protocols for both scenarios while AODV shows a better result in delay than DSR and TORA. The lowest value of network load is generated in both scenarios when TORA is used than when any of DSR and AODV is used. Thus, DSR because of its highest value of throughput outperformed others, followed by AODV and then TORA.

Keywords: AODV, DSR, MANETs, performance analysis, performance

II. INTRODUCTION

“Mobile Ad hoc Networks (MANETs) are fundamental element of pervasive networks, where user can communicate anywhere, any time and on-the-fly. MANETs introduce a new communication paradigm, which does not require a fixed infrastructure – they rely on wireless terminals for routing and transport services” [11]. A mobile ad hoc network is a self-configuring network of mobile devices or nodes connected by wireless links. It is a collection of mobile nodes that communicate together without any stationary infrastructure or predetermined topology of wireless links[1]. Every nodes in a mobile ad-hoc network can move freely in any direction and can change links to other nodes independently [2]. Each nodes dynamically discover other nodes such that, they can communicate together directly.

However, not all nodes can communicate directly to each other due to some signal transmission problems. Thus, each node function as a router and forward unrelated traffic to its own to others. The major challenge in MANET is maintaining the information needed to properly route traffic. Hence, mobile nodes are needed to forward packets on behalf of other nodes and deliver data through the network. An important characteristic of mobile ad hoc networks is changes in connectivity and link caused by node mobility and power requirement [1].

An ad hoc network is built around technology such as radio frequency (RF), global positioning system (GPS), Infrared (IR), Bluetooth and others. All the mobile devices or nodes in MANETs posses transmitters and receivers for communicating with one another [1]. MANET is suitable for many applications because of its flexibility and many other features such as easy set up and establishment of temporary communication in a disaster area without any pre-installed infrastructure [16].

II. RELATED WORK

Much work has been done in the area of routing protocols in MANETs. Different protocols had been evaluated using different kind of simulators most especially ns2. Some of these works are described next. Golan et al [3] worked on simulation-based comparative study of two On Demand Routing Protocols (DSR and TORA) using quantitative performance metrics such as data dropped, delay, throughput and media access delay for the analysis. The simulation result indicated that DSR performed better because of its less overhead control packets, and yielded more throughput than TORA. The usefulness of [3] is that, it help to know some of the metrics needed in evaluating the routing protocols. [4] worked on the design of scenario-based experiment and analyzed the performance of SEAD, which is a secure table driven routing protocol that is based on DSDV. [5] studied DSR, AODV, DSDV, CBRP and PAODV.
Their performance using different scenarios and workload were compared. His results show that CBRP has a higher overhead compared to DSR due to its periodic hello messages while AODV has the shortest end-to-end packet delay. DSR is once again a better consideration for this work.

In [6], three routing protocols AODV, DSR and DSDV were analyzed through three realistic scenarios. By using average traffic load, it was shown that DSR performed better than AODV when tested with the mobility values. This helps to know different scenarios that can be created to evaluate the performance of routing protocols.

In [7], several routing protocols were evaluated using packet-level simulations. The authors concluded that the on-demand routing protocols use lower routing load as compared to the traditional distance vectors and link state protocols. [8] Described the study of the performance evaluation of mobility speed on MANET. DSR, AODV and DSDV were compared using ns-2. Packet routing overhead, packet delivery ratio and normalized routing load were used as their performance metric. It was concluded in [8] that no winner emerge among the protocols used because different mobility patterns give different performance.

Although [9] tried to compare TORA, LDR and ZRP using OPNET; it could not compare TORA along the most common reactive protocols. Different scenarios were simulated in [10] using AODV and DSR as protocols as well as throughput, delay and receiving traffic as the performance metrics. OPNET 12 was used for this design with Campus of “Graphic Era University” as the realistic scenario. Also used is standard application (FTP) situated at the centre of Graphic Era University, 20 nodes with the two protocols enabled. The authors reported that in Wireless LAN throughput, AODV is slightly better than DSR. In the Wireless delay, it was reported that AODV is better.

Having done so much study on the previous related work, it was discovered that much have not been done in evaluating the performance of AODV, DSR and TORA together. Research effort have not focused much in evaluating their performance under a variable number of nodes. Moreover, most of the works described above used ns2 as a simulator whereas this work will be done in OPNET Modeler environment.

3. ROUTING PROTOCOLS

Routing in mobile ad hoc networks is quite different from conventional routing in wired networks. A dynamic routing protocol is needed for mobile ad hoc network to function properly in a rapidly changing network topology. Thus, routing here is really challenging because a node acts as both node and a router (or server). The Internet Engineering Task Force (IETF) through a MANET working group has published many of these protocols.

The purpose of this group is to standardize IP routing protocol that is suitable for wireless routing application within static and dynamic topologies [9]. Ad hoc routing protocols are grouped based on the network structure as flat routing, hierarchical routing and geographic assisted routing as shown below.

Figure 1: Classification of Routing Protocols in MANET

In flat routing, nodes communicate directly with each other and there is no clustering. It can be further classify into proactive, reactive and hybrid. Proactive protocols follow the strategies which are mostly followed by conventional routing protocols [9]. In proactive, routes are calculated before one is needed and up-to-date routing information is kept by all nodes every time. In reactive protocols, a route is only calculated when it is needed and does not keep routing information to all node every time. Also, a proactive scheme require a small delay in determining the route while in reactive scheme, a significant amount of delay is needed for creating a route. Hybrid protocols incorporate both properties of proactive and reactive protocols. Hierarchical routing protocols are used in a larger network where flat routings protocols are struggling with constraints [9]. They build and extend clusters and manage the communication inside a cluster. In Geographic position routing, no routing table is required, information can be send in any way in the direction of the destination. Also, no overhead is required to find or update routes, however position is required.

A. Ad Hoc On-Demand Distance Vector Protocol (AODV)

AODV is a reactive unicast routing protocol based on the distance vector algorithm. It allows mobile nodes to communicate, to form ad-hoc networks in a timely and self-configuring manner. It offers routes quickly and on demand and nodes do not need to maintain routes to destination nodes that are not communicating [6]. AODV also allows mobile nodes to adjust to network changes such as link outages. Although it is a distance vector protocol, based on the Bellman Ford Algorithm, it avoids the counting to infinity problem associated with distance vector protocols and ensures a loop-free routing.
Routing information in AODV is maintained at the active paths and routing tables at nodes. Each node here contains the routing information of the next hop. The routing table will expire if it is not used or reactivated for a specified time or period and the route will be broken. To discover a route, a route discovery operation is initiated and then send packets from a source node to a destination node. This route discovery operation consists of broadcasts route request (PREQ) packets, the broadcast ID, destination sequence number as well as the sequence number of the source node. A node in MANET sends hello messages to notify its existence to its neighbors or to discover its neighbors and monitor the links status to the next hop in active route. When a link disconnection occurs, a broadcast route error (RERR) packet is sent by a node to its neighbors, which then propagates the PERR packet towards other nodes that may be affected by the link [12]. A route discovery operation will now be reinitiated if the route is still needed.

B. Dynamic Source Routing Protocol (DSR)

DSR is a reactive routing protocol that uses source routing algorithm to pass data across the network. Each data packet header carries the sequence of nodes from which the packet must pass. This means that the intermediate nodes need only to keep track of their immediate neighbours to forward data packets while the source nodes need to know the entire hop sequence to the destination.

In DSR, the route acquisition procedure requests a route by flooding a Route Request packet similar to AODV. When a node receives a Route Request packet, it searches for all known routes to the requested destination in its route cache. When no route is found, it adds its own address to the hop sequence in the Route Request packet and forwards the packet further on. The Route Request packet then propagates through the network until it reaches a node with the route to the destination or the actual destination [5]. A Route Reply packet with the proper hop sequence for reaching the destination is then unicasted back to the source after finding the route. It should be noted that DSR does not rely on the bi-directional links since the Route Reply packet is sent to the source node through a route already stored in the route cached of the replying node, or by being piggybacked on a Route Request packet to the source node [5]. However, the reverse path in the Route Request can be used by the Route Reply message [5].

The advantage of DSR protocol is that, routes can be learnt from the source routes in the received packets. For example, a node ‘A’ can find a route to node ‘C’ through node ‘B’. In this way ‘A’ will learn route to ‘B’ and ‘C’ will also learn route to ‘A’ and so on [13]. This however leads to increase in flooding of the network with Route Request messages. To avoid this, the route acquisition procedure first queries the neighbouring nodes to see if they have a route to the intending destination. This can be done by sending a first Route Request message with zero hop-limit which will not be forwarded by neighbours. If there is no response from this initial request, then a new Route Request message will be flooded through the entire network.

DSR uses the MAC layer to tell nodes about link failures. When there is a link failure, a Route Error packet will be send back to the source node, which will then removes the broken link from its route cache as well as truncating all routes to the broken link. Moreover, an intermediate node that forwarded the Route Error will update its route cache in a similar way. Generally, DSR operates in two ways viz: Route discovery and Route maintenance. Both are on demand and does not require a periodic hello messages or link state advertisement between neighbours. DSR also supports multiple routes (thus large amount of overhead) to destinations because nodes use a caching property to maintain multiple routes. The routes can be learned during a route discovery or from other control packets they overheard (overhearing property). Multiple route support means that if one route fails, node using DSR can quickly pick another route from its cache. Caching also reduces overhead require in performing a new route discovery every time a route fails.

C. Temporarily-Ordered Routing Algorithm (TORA)

“The Temporally Ordered Routing Algorithm (TORA) is a highly adaptive, efficient and scalable distributed routing algorithm that is based on the concept of link reversal” [14]. TORA can be used for highly dynamic mobile and multi-hop networks. It is a source-initiated on-demand routing protocol that finds multiple routes from a source node to a destination node. Its main feature is that the control messages are typically localized to a very small set of nodes near the occurrence of a topological change [13][14]. To do this, routing information about adjacent or nearby node is maintained.

The network topology in TORA is regarded as a directed graph. A Directed Acyclic Graph (DAG) can be accomplished for this network by giving each node i a height hi [15]. A directional link from i to j means that hi is greater than hj (hi > hj). The height of the node defined as a quintuple includes the logical time of a link failure, a reflection indicator bit, propagation ordering parameter as well as a unique ID of the node. Packet moves from upstream downward according to the difference in height between nodes. DAG gives TORA the ability that many nodes can send packets to a given destination as well as guarantees that all routes are loop-free.

TORA can be separated into three main functions: route creation, route erasure and route maintenance. The route creation begins by setting the height of the destination to ‘0’ and heights of the remaining nodes to NULL (i.e. not defined). The source then broadcasts a QRY packet containing ID of the destination and a node with a not defined height responds by broadcasting a UPD packet containing the height of its own [15]. When a node receives a UPD packet, it sets its height to be greater than that of the UPD generator. A node with higher height is then considered as upstream and the node with a lower height as downstream. Hence, a DAG is constructed from the source to the destination and thus exist multiple paths route.
The DAG in TORA may be disconnected due to the node mobility. Route maintenance operation is therefore necessary in TORA. It has a unique characteristic that control messages are localized into a small set of nodes near the occurrence of topology changes. After losing its last downstream link, a node generates a new reference level which is broadcasted to its neighbours. Hence, links are reversed to reflect topology changes and adapt it to the new reference level. The route erasure operation in TORA floods CLR packets into the network and erase all invalid routes. The three protocols can be compared as follows:

Table 1: Comparing the three Routing Protocols

<table>
<thead>
<tr>
<th>Parameters</th>
<th>AODV</th>
<th>DSR</th>
<th>TORA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source Routing</td>
<td>No</td>
<td>Yes</td>
<td>NO</td>
</tr>
<tr>
<td>Topology</td>
<td>Full</td>
<td>Full</td>
<td>Reduced</td>
</tr>
<tr>
<td>Broadcast</td>
<td>Full</td>
<td>Full</td>
<td>Local</td>
</tr>
<tr>
<td>Update Information</td>
<td>Route error</td>
<td>Route error</td>
<td>Node’s height</td>
</tr>
<tr>
<td>Update destination</td>
<td>Source</td>
<td>Source</td>
<td>Neighbours</td>
</tr>
<tr>
<td>Method</td>
<td>Unicast</td>
<td>Unicast</td>
<td>Broadcast</td>
</tr>
</tbody>
</table>

4. EXPERIMENTAL SETUP

In this work, two scenarios have been created and analyzed for varying the number of nodes. The first scenario consists of 20 mobile nodes and it called Conference scenario. It has a low mobility compare to other one as only few of the people or nodes are moving at any point in time. Due to low mobility and low node density, the congestion here are low. The second scenario which made up of 40 mobile nodes is called Event scenario. This scenario models a group of 40 mobile people which are changing position more frequently than conference scenario. Thus they have high mobility, high node density and more congested. The purpose of these scenarios is to model a set of usage in a more realistic manner. The three performance metric used are discussed below.

A. Delay

Delay is refers to as the time it takes a packet to go from the source to the destination. It is expressed in seconds and thus called packet end-to-end delay. It is also refers to as latency. Because some applications are very sensitive to delay, they are often called delay sensitive applications.

B. Throughput

Throughput can be defined as the ratio of the total data that reaches a receiver from the sender. It is expressed as bytes or bits per second. Throughput can be affected by many factors such as limited bandwidth, network topology changes, and unreliable communication between nodes.

C. Network load

This is overall load of the network in bit per second submitted to wireless LAN lower layers by all higher layers in all nodes of the network. For a network to be efficient, it should be able to cope with large amount of traffic coming in. MANET routing packets are affected by high network loads; these increase the collisions of the control packet and eventually slow down the packet delivery in the channel.

D. Simulation Setup

One WLAN server was configured with FTP application in each scenario and three other important configurations for standard application were made. The size of the ftp file was set to 10000000 bytes and inters request time of 3600 seconds was used. The simulation environment size of 1000x1000 meters was also used. The configuration parameters used for mobility, ftp application, and profile are shown in table 2, 3 and 4 below.

Table 2: Mobility configuration
5. RESULT ANALYSIS AND OBSERVATIONS

The results of the simulations of the two scenarios and the observation made from the graph are discussed below. Appropriate conclusion on which protocol has a better performance in each scenario using the three metrics was made.

A. Analysis of AODV using the three metrics

The figure 2 shows how AODV can be compared in the two scenarios using delay as the metric. It can be noted that the delay in the conference scenario is lower than in the event scenario although the difference is not that significant. Since more nodes were involved in the event scenario, one can say that more hello messages will be sent to discover neighbour and more delay would be experienced. Hence for this work, AODV performs better in terms of delay in a network with fewer nodes (conference scenario) than in network with more nodes (event scenario). In figure 3, AODV is compared in the two scenarios using network load as a metric.

The result indicated that the value of network load generated is lower in the event scenario than in the conference scenario. This means that AODV perform well in terms of network load generated when more nodes are used than for fewer nodes. Also in figure 4, AODV performs very well in terms of throughput which means better performance for fewer nodes i.e. conference scenario than in event scenario with more nodes. In general, AODV shows a high performance in network with fewer nodes than the network with more nodes.
B. Analysis of DSR using the three metrics

Figure 5, 6 and 7 show how DSR can be compared using the two scenarios discussed in this work. In figure 5, the delay in conference scenario i.e. fewer nodes network is low compared to the event scenario with more nodes because the number of source packets or information forward to the next nodes is high in the event scenario and thus experience more delay. The network load generated in the conference scenario is almost the same as the one generated in the event scenario as shown in figure 6 except a little difference. In figure 7, throughput is higher in conference scenario with fewer nodes than in the event scenario with more nodes because in event scenario, more multiple routes are needed to be supported thereby increasing the overhead and subsequently deteriorating the network performance (i.e. reducing the throughput). Thus, for this work, DSR performs better generally in network with fewer nodes i.e. conference scenario that in the network with more nodes i.e. event scenario.

C. Analysis of TORA using the three metrics

Figure 8, 9, and 10 show how TORA is compared in the two scenarios using the three performance metrics considered in this work. The delay in figure 7 is very high for event scenario with more number of nodes than in conference scenario with less nodes. More time are needed to maintain multiple routes to destination in event scenario with more nodes than in conference scenario. Hence the delay graph is high. In figure 9, conference scenario with fewer nodes generated a high value of network load compared with the event scenario with more nodes. And in Figure 10, the throughput is higher in the conference scenario than in the event scenario because in the event scenario more control packets are transmitted to update the routing information immediately after a change in topology occur in event scenario than in conference scenario. This therefore lowers the throughput of the network with more nodes. Again, DSR performs generally better in a network of smaller nodes than the network with a higher number of nodes.
6. CONCLUSION

The aim of this study to evaluate the performance of three routing protocols under a variety of conditions has been successful with an appropriate conclusion drawn from the results and observations. Summary of the main points from the study are:

AODV and DSR outperform TORA in all scenarios. In both scenarios, DSR perform better than AODV and TORA in terms of throughput, with AODV showing a moderate result. AODV outperforms DSR and TORA by having the lowest value of delay in both scenarios, with DSR showing a moderate result. TORA perform well than both AODV and DSR in terms of network load generated in both scenarios. DSR shows a good performance than AODV in 20 nodes network i.e. conference scenario but cannot display the same good performance compared to AODV when the number of nodes is 40 i.e. in the event scenario. DSR begins to degrade in performance in terms of network load generated (this is due to source routes carried in every packet header) as the node density is increase in the network and according to the theoretical study of the DSR.

Since a high throughput is desirable of every network, and with DSR showing a higher throughput for these two scenarios under consideration, one can therefore concludes that DSR performs better than AODV and TORA while AODV performs better than TORA. However, the performance of these protocols may vary by varying the network and the parameters.
7. RECOMMENDATION
A more complex or realistic scenario could be carried out on the three protocols to be able to evaluate their performance better using a very large number of nodes especially to know the behaviour of DSR when nodes are more than 200. Since OPNET was used in this study, another simulator can be used to see how the three protocols are compared under the same performance metric and same number of nodes. Others can vary the parameters, scenarios and general model implementation. Performance metrics such as data drop, Packet Delivery Fraction (PDF) and Normalized Routing Load (NRL) can be used to further determine the best protocol.

The study of the performance of protocols could also be achieved using a real life Ad hoc network. This will be of utmost interest in the academic environment.

REFERENCES

Multi-Agent Based Patient Scheduling Using Ant Colony Optimization

A. A. Obiniyi  
Department of Mathematics  
Ahmadu Bello University  
Zaria, Nigeria

M. K. Ahmed  
Department of Mathematics  
Gombe State University  
Gombe, Nigeria  
E-Mail: kabirmka@yahoo.com  
Phone Number: +2348039285289

ABSTRACT

With the advent of Agent Technology, and challenges facing the health care sector as a result of scarcity of both medical personnel and resources and also the distributive and complex nature of the activities in the medical domain, there is need to use the new programming paradigm in order to come up with robust and reliable system that will solve problems bedeviling the health care sector especially in the developing countries like Nigeria. In this work, a Meta heuristic technique called Ant Colony Optimization (ACO) was used in the patient scheduling in order to find the best possible solutions with small computational efforts that will reduce the waiting time of patients when competing for the various medical resources. Our ACO approach has its waiting time more than previously studied Particle Swarm Optimization (another Meta heuristic technique).

Keyword: Agent, Pheromone, ACO

1. INTRODUCTION

The hospitals contain different autonomous wards and ancillary units. Each department has the authority to take their own decision so it is decentralized. In addition to the complexity arising from the distributed structure of hospitals, patient scheduling has to be performed in the face of a high degree of uncertainty about the treatment pathways of patients within the hospital [1]. Hence the need to get an optimal patient scheduling using Meta heuristic technique that will improve the utilization of Medical Personnel and other health care resources due their scarcity that enable patient to book for appointment or service at their convenience.

The patients arrive continuously at the hospital and the necessary medical treatments are often not able to be completely determined at the beginning of the treatment process. Moreover the results of a diagnostic examination might change the (medical) priority of the patients, invoke additional activities and/or make other medical actions obsolete [2], [3]. The aim of this research is to find an optimal Patient schedule using meta-heuristic technique Ant Colony Optimization.

Ant colony optimization (ACO) takes inspiration from the foraging behavior of some ant species. These ants deposit pheromone on the ground in order to mark some favorable path that should be followed by other members of the colony. Ant Colony Optimization exploits a similar mechanism for solving optimization problems. The behavior of each ant in nature is that it wanders randomly at first, laying down a pheromone trail, if food is found, then return to the nest laying down a pheromone trail, if pheromone is found, with some increased probability follow the pheromone trail, once back at the nest, go out again in search of food. However, pheromones evaporate over time, such that unless they are reinforced by more ants, the pheromones will disappear. [5]

2. RELATED WORK

Ant colonies, and more generally social insect societies, are distributed systems that, in spite of the simplicity of their individuals, present a highly structured social organization. As a result of this organization, ant colonies can accomplish complex tasks that in some cases far exceed the individual capabilities of a single ant. The field of “ant algorithms” studies models derived from the observation of real ants’ behavior, and uses these models as a source of inspiration for the design of novel algorithms for the solution of optimization and distributed control problems. [4]
In [6] they find a feasible schedule for a given task set to a set of heterogeneous processors without exceeding the capacity of the processors, which is NP-Hard. The study uses a paradigm using Ant Colony Optimization (ACO) for arriving at a schedule. An attempt is made to arrive at a feasible schedule of a task set on heterogeneous processors ensuring load balancing across the processors. The heterogeneity of the processors is modeled by assuming different utilization times for the same task on different processors. The two parameters: average wait time of tasks and utilization of processors are computed using the First Come First Served (FCFS). This approach to the tasks assignment problem using ACO performs better with respect to the two parameters used compared to the FCFS algorithm but the time taken to come up with the schedule using ACO is slightly more than that of FCFS.

In a similar development [8] shows that the problem of determining a set of periodic tasks can be assigned to a set of heterogeneous processors in such a way that all timing constraints are met, in general, to be NP-hard. But in this case the authors used MAX-MIN Ant System (MMAS) for the update of the pheromone trials. The paper presents a new algorithm based on Ant Colony Optimization (ACO) Meta heuristic for solving this problem. The ACO approach to the task assignment problem is presented. Preliminary test shows that this approach has better performance than a GA heuristic and Baruah’s approximation algorithm In addition to being able to search for a feasible assignment solution; the approach can further optimize the solution to reduce its energy consumption.

In this research MAX-MIN Ant System will also be used in the update of the pheromone because only one single ant is used to update the pheromone trails after each iteration. In addition to that, since the key to achieve best performance of ACO algorithms is to combine an improved exploitation of the best solutions found during the search with an effective mechanism for avoiding early search stagnation. MAX–MIN Ant System has been specifically developed to meet these requirements. [9]

From [7] Healthcare organizations are facing the challenge of delivering high-quality services through effective process management at all levels-locally, regionally, nationally, and internationally such as Patient scheduling. The researcher present an agent based information services for mobile users. Multi Agent systems (MAS) have great application in medical domains, due to the characteristics of the problems in this area and are the basis of an emerging technology that promises to make it much easier to design and implement. The authors identify problems of meeting scheduling (MS) in the health care sectors. In which they identify the need to implement an agent based meeting scheduling system, which can schedule a meeting between Patients and Doctors.

The main contribution of the research was creation of an automated meeting scheduling agent in health care domain that can serve the following functions: It allows the patients to input his/her meeting request, Negotiates with the agents of the other requested patients, Finds out its best fitting and free time slots, Compares them with the sent fitting slots of the patient-agents and find out the best ones, Reacts to the incoming patient request by sending back its best fitting free time slots, Shows all fixed meeting in a time table and allows the patient to input his/her preferences.

In [10] the paper focuses on finding an optimal schedule using the meta-heuristic technique Particle Swarm Optimization (PSO) and coordinating the hospital environment using multi-agents. Since agents are suitable in resource allocation because of their coordination and social abilities. This agent based approach is implemented in JADE (Java Agent Development Environment) and tested for different data sets. They compared the results with the traditional dispatching rules like First Come First Serve (FCFS), Minimum Slack (MS), Shortest Processing Time (SPT) and Longest Processing Time (LPT).

The authors were motivated by the real needs in hospital environments; the paper focuses on finding an optimal schedule using the meta-heuristic technique Particle Swarm Optimization (PSO) and coordinating the hospital environment using multi-agents. The basic framework is implemented in the JADE platform. JADE (Java Agent Development Framework) is a software framework fully implemented in Java language in which the final schedule has been generated. There is a considerable percentage of reduction in the number of late patients in PSO method when compared to other scheduling techniques except SPT. The results clearly shows that the performance improvement of total weighted waiting time in APS-PSO is 4.13% when compared to the best performing dispatching rule MS and 52% when compared to LPT.

3. AGENT BASED PATIENT SCHEDULING WITH ANT COLONY OPTIMIZATION

The framework consists of Patient Agent (PA), Resource Agent (RA), Common Agent (CA) and Main Agent (MA). The CA is a general physician. According to the symptoms given by the patient the CA determines the initial treatments that the patient has to undergo. The treatment consists of consultation of Specialist doctor, diagnostic procedures like X-Ray, lab tests etc. Now the PA knows what are tasks it has to perform and then it request for the resources. The resource agent may be X-Ray, Lab Tests, consultation with Specialist doctor etc. Each RA has multiple time slots. The PA requests for this time slots. Multiple PA may request for the same slot. For each resource a RA is created. In addition to these agents an ACO agent is used to perform the ACO optimization. On arrival of the patients, this ACO agent is called to perform optimization using ACO and an optimized schedule is generated.
4. ANT COLONY OPTIMIZATION FOR PATIENT SCHEDULING

Patient Scheduling is the process of scheduling and sequencing the patient for various multiple resources in the healthcare domain. The multiple constraint and goals to be achieved in minimal time this problem highly complex. In the previous work (Particle Swarm Optimization) they get the value of Min and Max from one of the experiment related to that optimization.

In this research we followed the MAX-MIN Ant System (MMAS) in for the update of pheromone trials. Regarding pheromone trial limits, an estimate of the upper bound, f (sbest ) / \( \rho \), is used to define \( \tau_{\text{max}} \), where sbest is the best-so-far solution, and \( \rho \) is the evaporation rate of pheromone trails. \( \tau_{\min} \) is defined to be \( \tau \).

For a patient scheduling problem (PSP) consisting of \( n \) patient (or jobs) \( P = \{P_1, \ldots, P_n\} \) and \( m \) resources \( R = \{R_1, \ldots, R_m\} \), each patient \( P_i \) has a set of sequential task (operations) \( O_{ij} \), \( i=1,2,\ldots,n; j=1,\ldots,n_i \leq m \); where \( i \) is index of the patient and \( j \) is the index of the task or operation.

\[
\tau_{\min} = \frac{\tau_{\max} \cdot (1 - \sqrt[n]{P_{\text{best}}^{\min}})}{(\text{avg} - 1) \cdot \sqrt[n]{P_{\text{best}}^{\max}}} \quad \cdots \quad 1[9]
\]

\[
\tau_{\min} = \frac{1}{1 - \rho} \cdot (1 - \sqrt[n]{P_{\text{best}}^{\min}}) \quad \cdots \quad 2
\]

Since,

\[
\tau_{\max} = \frac{1}{1 - \rho} \quad \cdots \quad 3 [9]
\]

Where \( P_{\text{best}}^{\min} \) = the probability of constructing best solution found.

\( n \) = the total number of items.

avg = average number of items to choose at every decision point when building a solution.

\( \rho \) = the evaporation rate of the pheromone. It value is set to 0.7 based on the study made in a related literatures.

\[
\tau_{\max} = 1/0.7
\]

= 3.33

To get \( P_{\text{best}}^{\min} \), we have to get 3P3 = 6, \( P_{\text{best}}^{\min} = 1/6 \)

\[
\tau_{\min} = \frac{3.33 \cdot (1 - \sqrt[6]{0.166})}{(2 - 1) \cdot \sqrt[6]{0.166}} \approx 2.725
\]

\[
\tau_{\min} = 2.725 \approx 2.73
\]

\[
\tau_{\max} = 3.333 \approx 3.33
\]

5. INITIALIZATION

An example of patient scheduling problem (PSP) for 3 patient and 3 resources are given in the table 1 below. The example initialization of pheromone update by the ant of the given PSP is described in table 2. The pheromone update is initialized with the random numbers between \( \tau_{\min} \) and \( \tau_{\max} \).

<table>
<thead>
<tr>
<th>Patient</th>
<th>Arrival Time</th>
<th>Task Sequence</th>
<th>Processing Time</th>
<th>Due Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>R1 R2 R3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>R3 R2 R1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>R1 R3 R2</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

| Task | 1 | 1 | 1 | 2 | 3 | 2 | 3 | 3 |

| Pheromone update | X | X | X | X | X | X | X | X |

| Values | 3.21 | 2.75 | 2.99 | 2.86 | 2.78 | 3.30 | 2.96 | 2.81 |

6. DECODING ANT INTO SOLUTIONS

For decoding ant into a schedule, the first step is to sort the values of the pheromone update vector into ascending order. The next step is to arrange the tasks in the corresponding order as the values of the position vector, sorted in step 1. The resultant sequence of task and the corresponding position values are given in table 3.
Table 3: An Example of Sequenced Ant

<table>
<thead>
<tr>
<th>Task</th>
<th>1</th>
<th>3</th>
<th>2</th>
<th>3</th>
<th>2</th>
<th>3</th>
<th>1</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pheromone Update</td>
<td>Xi2</td>
<td>Xi9</td>
<td>Xi5</td>
<td>Xi8</td>
<td>Xi4</td>
<td>Xi7</td>
<td>Xi3</td>
<td>Xi1</td>
<td>Xi6</td>
</tr>
<tr>
<td></td>
<td>2.75</td>
<td>2.76</td>
<td>2.78</td>
<td>2.81</td>
<td>2.86</td>
<td>2.96</td>
<td>2.99</td>
<td>3.21</td>
<td>3.30</td>
</tr>
</tbody>
</table>

Thus from the above Table 3, the operational based permutation \( \pi = (1, 3, 2, 3, 2, 3, 1, 1, 2) \) is obtained. An element of \( \pi \) with \( i \) value stands for \( P_i \). The \( j \) th occurrence of \( i \) in \( \pi \) refers to \( O_{ij} \) that is the \( j \) th task (operation) of patient. The precedence of the task is determined simply by the order of the elements of \( \pi \). Considering the precedence constraint of the task is \( O_{11}, O_{31}, O_{21}, O_{32}, O_{22}, O_{33}, O_{12}, O_{13}, O_{23} \). Ordered tasks from \( \pi \) are ready to be scheduled. The table 4 below gives the Decoded Schedule using ACO.

7. RESULT AND DISCUSSION

Consider the operation-based permutation \( \pi = (1, 3, 2, 3, 2, 3, 1, 1, 2) \), the first element in the permutation is 1; therefore the 1st task of the 1st patient is processed on resource R1. In succession, the 2nd element in the permutation is 3, hence the 1st task of the 3rd patient is processed on R1, and then the 3rd element is 2, so the 1st task of the 2nd patient is processed on R3. Thus based on the above description the decode schedule in Table 4 is obtained.

Table 4: Decoded Schedule using ACO

<table>
<thead>
<tr>
<th>R3</th>
<th>P2</th>
<th>P3</th>
<th>P1</th>
</tr>
</thead>
<tbody>
<tr>
<td>R2</td>
<td>P2</td>
<td>P3</td>
<td>P1</td>
</tr>
<tr>
<td>R1</td>
<td>P1</td>
<td>P3</td>
<td>P2</td>
</tr>
<tr>
<td>Timeslot</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>
So when we compare the patient scheduling using ant colony optimization with the conventional first come first serve (FCFS). The waiting time in FCFS is less than that of ACO. While in Particle Swarm Optimization (PSO) is less than that of FCFS as shown in the figure 5 and 6.

**Table 5: Decoded Schedule using FCFS**

<table>
<thead>
<tr>
<th>R3</th>
<th>P2</th>
<th>P3</th>
<th>P1</th>
</tr>
</thead>
<tbody>
<tr>
<td>R2</td>
<td>P1</td>
<td>P2</td>
<td>P3</td>
</tr>
<tr>
<td>R1</td>
<td>P1</td>
<td>P3</td>
<td>P2</td>
</tr>
</tbody>
</table>

| Timeslot | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |

**Table 6: Decoded Schedule of Particle Swarm Optimization. [10]**

<table>
<thead>
<tr>
<th>R3</th>
<th>P2</th>
<th>P3</th>
<th>P1</th>
</tr>
</thead>
<tbody>
<tr>
<td>R2</td>
<td>P1</td>
<td>P2</td>
<td>P3</td>
</tr>
<tr>
<td>R1</td>
<td>P3</td>
<td>P1</td>
<td>P2</td>
</tr>
</tbody>
</table>

| Timeslot | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |

**8. CONCLUSION**

In this paper, an agent-based patient scheduling using ACO has been explained and presented. Even though ACO performs better than other algorithms when it comes to task assignment problem in multi-processor environment but when it comes to patient scheduling it is never the same. From the result it is clearly shown that the total waiting time in ACO is nine (9) units which is more than that of a FCFS and PSO that is six (6) and four (4) Unit of the time slot respectively.
Reference


A New Approach: Automatically Identify Proper Noun from Bengali Sentence for Universal Networking language

M. S. Islam & J.K. Das
Computer Science and Engineering
Jahangirnagar University
Savar, Dhaka-1213, Bangladesh
Phone: +8801916574623, email: syefulislam@yahoo.com
Phone: +8801712509082, email: drdas64@yahoo.com

ABSTRACT

More than hundreds of millions of people of almost all levels of education and attitudes from different country communicate with each other for different using various languages. Machine translation is highly demanding due to increasing the usage of web based Communication. One of the major problem of Bengali translation is identified a naming word from a sentence, which is relatively simple in English language, because such entities start with a capital letter. In Bangla we do not have concept of small or capital letters and there is huge no. of different naming entity available in Bangla. Thus we find difficulties in understanding whether a word is a proper noun or not. Here we have introduce a new approach to identify proper noun from a Bengali sentence for UNL without storing huge no. of naming entity in word dictionary. The goal is to make possible Bangla sentence conversion to UNL and vice versa with minimal storing word in dictionary.

Keywords- UNL; Rule based analysis, Morphological Analysis; Post Converted; Head word; Knowledge base.

I. INTRODUCTION

Today the demand of inter communication between all levels of peoples in various country is highly increased. This globalization trend evokes for a homogeneous platform so that each member of the platform can apprehend what other intimates and perpetuates the discussion in a mellifluous way. However the barriers of languages throughout the world are continuously obviating the whole world from congregating into a single domain of sharing knowledge and information. Therefore researcher works on various languages and tries to give a platform where multi lingual people can communicate through their native language. Researcher analyze the language structure and form structural grammar and rules which used to translate one language to other. From the very beginning the Indian linguist Panini proposed vyakaran (a set of rules by which the language is analyzed) and gives the structure for Sanskrit language. After the era of Panini various linguist works on language and proposed various technique.

But the most modern theory proposed by the American linguist Noam Chomsky is universal grammar which is the base of modern language translation program. From the last few years several language-specific translation systems have been proposed. Since these systems are based on specific source and target languages, these have their own limitations. As a consequence United Nations University/Institute of Advanced Studies (UNU/IAS) were decided to develop an inter-language translation program [1].

The corollary of their continuous research leads a common form of languages known as Universal Networking Language (UNL) and introduces UNL system. UNL system is an initiative to overcome the problem of language pairs in automated translation. UNL is an artificial language that is based on Interlingua approach. UNL acts as an intermediate form computer semantic language whereby any text written in a particular language is converted to text of any other forms of languages [2]-[3].

UNL system consists of major three components: language resources, software for processing language resources (parser) and supporting tools for maintaining and operating language processing software or developing language resources. The parser of UNL system take input sentence and start parsing based on rules and convert it into corresponding universal word from word dictionary. The challenge in detection of named is that such expressions are hard to analyze using UNL because they belong to the open class of expressions, i.e., there is an infinite variety and new expressions are constantly being invented. Bengali is the seventh popular language in the world, second in India and the national language of Bangladesh. So this is an important problem since search queries on UNL dictionary for proper nouns while all proper nouns (names) cannot be exhaustively maintained in the dictionary for automatic identification.
This work aims at attacking this problem for Bangla language, especially on the human names detection from Bengali sentence without storing naming word in dictionary.

This research paper is organized as follows: Section 2 deals with the problem domain and Section 3 provides the theoretical analysis of The Universal Networking Language. In section 4 the functioning of En-Converter is described. Section 5 Introduce the new approach to identify proper noun from a Bengali sentence for UNL. Section 6 gives the results corresponding to analysis. Finally Section 7 draws conclusions with some remarks on future works.

2. PROBLEM DOMAIN

UNL is an electronic language for computers. It intermediates understanding among different natural languages. UNL represent sentences in the form of logical expressions, without ambiguity. These expressions are not for humans to read, but for computers. It would be hard for users to understand, and they would not need to, unless they are UNL experts. Thus, UNL is an intermediate language to be used through the internet, which allows communication among people of different languages using their mother tongue. Adding UNL to the network platforms will change the existing communication landscape. The purpose of introducing UNL in communication network is to achieve accurate exchange of information between different languages. Information has to be readable and understandable by users. Information expressed in UNL can be converted into the user’s native language with higher quality and fewer mistakes than the computer translation systems. In addition, UNL, unlike natural language, is free from ambiguities. Currently, the UNL includes 16 languages, which are the six official languages of the United Nations (Arabic, Chinese, English, French, Russian and Spanish), in addition to the ten other widely spoken languages (German, Hindi, Italian, Indonesian, Japanese, Latvian, Mongol, Portuguese, Swahili and Thai).

Researchers already start works on Bengali language to include it with UNL system. Human language like Bangla is very rich in inflections, vibhakties (suffix) and karakas, and often they are ambiguous also. That is why Bangla parsing task becomes very difficult. At the same time, it is not easy to provide necessary semantic, pragmatic and world knowledge that we humans often use while we parse and understand various Bangla sentences. Bangla consists of total eighty-nine part-of-speech tags. Bangla grammatical structure generally follows the structure: subject-object-verb (S-O-V) structure [4-5]. We also get useful parts of speech (POS) information from various inflections at morphological parsing. But the major problem is identifying the name from sentence and convert is very difficult. In this section we try to clear the problem domain and define some point why the processing of naming word is difficult.

In terms of native speakers, Bengali is the seventh most spoken language in the world, second in India and the national language of Bangladesh. Under the project of UNL society the Bengali linguist works on Bangla language. They already introduce some rules for UNL.

Named identification in other languages in general but Bengali in particular is difficult and challenging as:
- There is huge no. of naming word available in Bangla and it’s not wise decision to store all of naming word in word dictionary. It causes slow performance.
- Unlike English and most of the European languages, Bengali lacks capitalization information.
- Bengali person names are more diverse compared to the other languages and a lot of these words can be found in the dictionary with some other specific meanings.
- Bengali is a highly inflectional language providing one of the richest and most challenging sets of linguistic and statistical features resulting in long and complex word forms.
- Bengali is a relatively free order language.

In Bengali language conversion, En-Converter parse the sentence word by word and find word from dictionary and apply rules. When En-Converter doesn’t find any word from dictionary then En-Converter creates a temporary entry for this word. In the maximum case the temporary entry is name word. We apply some rules to ensure that this words which is not in dictionary (temporary entry) are proper noun.

The later sections we discuss about a new technique of identifies the proper noun from Bangla sentence and defines a post converter for convert the Bangla name to universal word.

3. THE UNIVERSAL NETWORKING LANGUAGE

The Internet has emerged as the global information infrastructure, revolutionizing access to information, as well as the speed by which it is transmitted and received. With the technology of electronic mail, for example, people may communicate rapidly over long distances. Not all users, however, can use their own language for communication.

The Universal Networking Language (UNL) is an artificial language in the form of semantic network for computers to express and exchange every kind of information. Since the advent of computers, researchers around the world have worked towards developing a system that would overcome language barriers. While lots of different systems have been developed by various organizations, each has their special representation of a given language. This results in incompatibilities between systems. Then, it is impossible to break language barriers in all over the world, even if we get together all the results in one system.
Against this backdrop, the concept of UNL as a common language for all computer systems was born. With the approach of UNL, the results of the past research and development can be applied to the present development, and make the infrastructure of future research and development.

The UNL consists of Universal words (UWs), Relations, Attributes, and UNL Knowledge Base. The Universal words constitute the vocabulary of the UNL, Relations and attribute constitutes the syntax of the UNL and UNL Knowledge Base constitutes the semantics of the UNL. The UNL expresses information or knowledge in the form of semantic network with hyper-node. UNL semantic network is made up of a set of binary relations, each binary relation is composed of a relation and two UWs that old the relation [6].

4. FUNCTION OF UNL EN-CONVERTER

To convert Bangla sentences into UNL form, we use En-Converter (EnCo), a universal converter system provided by the UNL project. The EnCo is a language independent parser, which provides a framework for morphological, syntactic and semantic analysis synchronously. Natural Language texts are analyzed sentence by sentence by using a knowledge rich lexicon and by interpreting the analysis rules. En-Converter generates UNL expressions from sentences (or lists of words of sentences) of a native language by applying En-conversion rules. In addition to the fundamental function of En-conversion, it checks the formats of rules, and outputs massages for any errors. It also outputs the information required for each stage of En-conversion in different levels. With these facilities, a rule developer can easily develop and improve rules by using En-Converter [7].

First, En-Converter converts rules from text format into binary format, or loads the binary format En-conversion rules. The rule checker works while converting rules. Once the binary format rules are made, they are stored automatically and can be used directly the next time without conversion. It is possible to choose to convert new text format rules or to use existing binary format rules.

- Convert or load the rules.
- Secondely, En-Converter inputs a string or a list of morphemes / words of a sentence s native language.
- Input a sentence.
- Then it starts a apply rules to the Node-list from the initial state (Fig. 1).

En-Converter applies En-conversion rules to the Node-list its windows. The process of rule application is to find a suitable rule and to take actions or operate on the Node-list in order to create a syntactic tree and UNL network using the nodes in the Analysis windows. If a string appears in the window, the system will retrieve the word dictionary and apply the rule to the candidates of word entries. If a word satisfies the conditions required for the window of a rule, this word is selected and the rule application succeeds. This process will be continued until tree and UNL network are completed and only the entry node remains in the Node-list.

- Apply the rules and retrieve the Word Dictionary.
- Output the UNL expressions.

Finally the UNL network (Node-net) is outputted to the output file in the binary relation format of UNL expression.

4.1 Temporary Entries

In the following two cases, En-Converter creates a temporary entry by assigning the attribute "TEMP"("TEMP" is the abbreviation for "Temporary").

- Except for an Arabic numeral or a blank space, if En-Converter cannot retrieve any dictionary entry from the Word Dictionary for the rest of the character string, or
- When a rule requiring the attribute "TEMP" is applied to the rest of the character string.

The temporary entry has the following format and it's UW, shown inside the double quotation "and", is assign to be the same as its headword (HW).
The next section we proposed the technique of identifies the proper noun from Bangla sentence and defines a post converter for convert the Bangla name to universal word.

5. AUTOMATICALLY PROPER NOUN IDENTIFICATION APPROACH

The proper noun conversion is relatively simple in English language, because such entities start with a capital letter. In Bangla we do not have concept of small or capital letters. Thus we find difficulties in understanding whether a word is a proper noun or not. For example, the Bangla word "BISWAS" can be a proper noun (i.e., a family name of a person) as well as an abstract noun (with the meaning of faith in English). For example, in order to understand the following Bangla sentence, we must need an intelligent parser. A parser takes the Bangla sentence as input and parses every sentence according to various rules [8-9].

Here we proposed a method for UNL to proper noun conversion which is a combination of dictionary-based, rule-based approaches. In this approach, UNL En-Converter identifies the proper noun using rules and morphological analysis. The approaches are sequentially described here and demonstrated in Fig.2.

5.1.1 Dictionary based analysis for proper noun detection

If a dictionary is maintained where we try to attach most commonly used proper noun and it is known as Name dictionary. Here we describe the dictionary based proper noun detection technique sequentially. Firstly the input sentence is processed on en-converter which finds the word on word dictionary. If the word is found then it is converted into relative universal word. If not in dictionary then En-Converter creates a temporary entry for this word. Secondly the en-converter finds the word with flag TEMP into name dictionary. If it is found then it is conceded as the word may be noun and apply rules to ensure. Finally if the word is not in name dictionary then it sends into morphological analyzer to conform that the word is proper noun.

5.1.2 Rule-based analysis for proper noun detection

Rule-based approaches rely on some rules, one or more of which is to be satisfied by the test word. Some words which use in Bangla sentence as a part of name. Here we take a technique to identify proper noun using such word (part of name). Firstly we make dictionary entry with pof (part of name) and other relevant attribute. To identify proper noun from Bangla sentence use pof word some typical rules are given below.
Rule 1- If the parser find the following word like (মান, মূল, মোদামাত, মোদামাত, আঃ, আন্দুল, আন্দুল, সিং, সিং, বিমান, বিমান, বিবিইটি, ডাক, ডাক, জন, জন, স্বামী, স্বামী, রোজা, রোজা, পীরুজ, পীরুজ),
then the word is considered as the first word of name and set the status of the word first part of name (FPN).

Rule 2- If the parser find the following word like (আবাস, আবাস, প্রেম, প্রেম, সীমা, সীমা, নায়ক, নায়ক, জাতি, জাতি, কোল, কোল, মুখ, মুখ, নর, নর, বিবিইটি, ডাক, ডাক, জন, জন, স্বামী, স্বামী),
then the word is with 1st, 2nd, 5th and 6th verbal inflexions (Bibhakti). So when parser fined an
unknown word with 1st, 2nd, 5th and 6th bibhakti then the
word may be proper noun [11-12].

Rule 3- From sentence structure if parser fined an unknown
word is in the position of kari karak and word is ended with
1st bibhakti then it is conceder as a proper noun.

Rule 4- If any sentence contains more than one word ended
with 1st bibhakti then the first word is with flag unknown
word then must be kari karak and the word is proper noun.

Rule 5- In case of apaadaan karak, if any word in the
sentence ended with 5th bibhakti and the word is unknown
then it must be noun. If most of all other’s noun are in word
dictionary so unknown word ended with 5th bibhakti must be
proper noun.

Rule 6- In case of adhikaran karak, if any word in the
sentence ended with 7th bibhakti and the word is unknown
then it may be the name of place. If the word is not in
dictionary then it is conceder as proper noun (name of any
pace).

5.1.3 Morphological Analysis for proper noun detection
When previous two steps fail to identify proper noun or there
is confusion about the word is proper noun or not then we
apply morphological analysis to sure that the unknown word is
proper noun. In this case we conceder the structure of words
and the position of word in the sentence and identify the word
type [10].
In that time, when En-Converter identify an word or collection of word as a proper noun and En-Converter convert into UNL expression it keep track temporary word using custom UNL relation tpl and tpr. We use two relation “tpr” and “tpl” to identify the word which should converted. The relation “tpr” use when En-Converter finds the temporary word after pof (part of name) attribute and “tpr” use when En-Converter finds the temporary word before pof (part of name ) attribute. When En-Converter found “tpr” relation then it converts the word which is after blank space. For the case of “tpl” it converts the first word. If proper noun contains only one word with attribute TEMP then using this TEMP attribute converter convert.

5.2 Function of Post Converter

In previous section we identify proper noun from bangle sentence and convert the sentence into corresponding intermediate UNL expression. But there is little bit problem, En-Converter convert those word which is in word dictionary. In the case of temporary word which is already identified as a proper noun or part of proper noun cannot converted and it is same as in Bengali sentence. It is difficult to other people who cannot read bangle language. So it must be converted into English for UNL expression. Post converters do this conversion. The function of post converter demonstrated in Fig. 5.

Step 2: In second step post converter read the full expression and fined relation tpr or tpl. The relation tpr and tpl are used to identify the word which should convert.

Step 3: At third step Post converter collect Bangla word which are converted within this post converter using the help of above two relations. When En-Converter found ”tpr” relation then it collects the word which is after blank space and for tpl it collects the first word.

Step 4: In this steps converter convert the word applying rules and finding the corresponding English syllable or word form syllable dictionary.

Step 5: In this step we get the converted word which is placed on final UNL expression.

Step 6: This steps is the final steps which generate final UNL expression.

Here we have listed some dictionary entries for post converter. Table 1 shows the Bengali vowel and table 2 shows the shows the Bengali consonant and the corresponding entries in dictionary. In table 3 it shows some dictionary entries for consonant plus vowel (kar). Here we only try to present how post converter converts Bengali to English. In future we define the full phonetics for Bengali to English conversion.

Table1. Dictionary entries for Bengali vowel

<table>
<thead>
<tr>
<th>Bangla vowel</th>
<th>Dictionary entries</th>
</tr>
</thead>
<tbody>
<tr>
<td>আ</td>
<td>[আ] &quot;a&quot; (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>আঁ</td>
<td>[আঁ] &quot;a&quot; (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>ই</td>
<td>[ই] &quot;i&quot; (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>ইঁ</td>
<td>[ইঁ] &quot;ei&quot; (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>ই঄</td>
<td>[ই঄] &quot;oo&quot; (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>ইঅ</td>
<td>[ইঅ] &quot;u&quot; (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>র</td>
<td>[র] &quot;rii&quot; (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>রঁ</td>
<td>[রঁ] &quot;a&quot; (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>রঃ</td>
<td>[রঃ] &quot;oi&quot; (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>রঅ</td>
<td>[রঅ] &quot;ou&quot; (TEMP) &lt;.,0,0&gt;</td>
</tr>
</tbody>
</table>

Figure. 5. Function of Post Converter

5.3 Proposed En-conversion process

Here I use simple phonetic bangle to English translation method. We use same En-Converter which is used in UNL. Firstly need to create dictionary for Bengali to English conversion. Then rules are define for converter which uses these rules for conversion. When En-Converter found “tpr” relation then it converts the word which is after blank space. For the case of “tpl” it converts the first word. The functions of post converter are sequentially described here and architecture of Post converter demonstrated in Fig.5.

Figure.5. Architecture of Post converter (Bengali to English)

5.3.1 Algorithm

How post converter converts Bangla word for intermediate UNL expression to English for final UNL expression. The process are describe step by step- Step 1: In first step the UNL expression is the inputs of post converter for find the final UNL expression.
Table 2. Dictionary Entries for Bengali Consonant

<table>
<thead>
<tr>
<th>Bangla consonant</th>
<th>Dictionary Entries</th>
</tr>
</thead>
<tbody>
<tr>
<td>ক</td>
<td>[ক] । “k” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>ধ</td>
<td>[ধ] ] “dh” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>ঘ</td>
<td>[ঘ] ] “gh” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>ঙ</td>
<td>[ঙ] ] “ng” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>চ</td>
<td>[চ] ] “ch” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>ছ</td>
<td>[ছ] ] “ch” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>জ</td>
<td>[জ] ] “ji” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>ঝ</td>
<td>[ঝ] ] “jh” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>ঞ</td>
<td>[ঞ] ] “niya” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>ট</td>
<td>[ট] ] “t” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>ঠ</td>
<td>[ঠ] ] “th” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>ড</td>
<td>[ড] ] “d” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>ঢ</td>
<td>[ঢ] ] “dh” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>ণ</td>
<td>[ণ] ] “n” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>ত</td>
<td>[ত] ] “t” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>থ</td>
<td>[থ] ] “th” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>দ</td>
<td>[দ] ] “d” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>ধ</td>
<td>[ধ] ] “dh” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>ন</td>
<td>[ন] ] “n” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>প</td>
<td>[প] ] “p” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>ফ</td>
<td>[ফ] ] “f” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>ব</td>
<td>[ব] ] “v” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>ভ</td>
<td>[ভ] ] “bh” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>ম</td>
<td>[ম] ] “mm” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>ন</td>
<td>[ন] ] “n” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>ল</td>
<td>[ল] ] “l” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>শ</td>
<td>[শ] ] “sh” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>ষ</td>
<td>[ষ] ] “s” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>স</td>
<td>[স] ] “s” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>হ</td>
<td>[হ] ] “h” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>ড়</td>
<td>[ড়] ] “r” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>ঢ়</td>
<td>[ঢ়] ] “rh” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>ণঃ</td>
<td>[ণঃ] ] “y” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>তঃ</td>
<td>[তঃ] ] “r” (TEMP) &lt;.,0,0&gt;</td>
</tr>
</tbody>
</table>

Table 3. Dictionary Entries for Bengali Consonant plus Bengali Kar

<table>
<thead>
<tr>
<th>Bangla parts of word</th>
<th>Dictionary entries</th>
</tr>
</thead>
<tbody>
<tr>
<td>কা</td>
<td>[কা] ] “ka” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>কি</td>
<td>[কি] ] “ki” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>কে</td>
<td>[কে] ] “ke” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>কঁ</td>
<td>[কঁ] ] “ko” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>কু</td>
<td>[কু] ] “ku” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>কৃ</td>
<td>[কৃ] ] “kri” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>কো</td>
<td>[কো] ] “kou” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>কা</td>
<td>[কা] ] “ka” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>কি</td>
<td>[কি] ] “ki” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>কে</td>
<td>[কে] ] “ke” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>কঁ</td>
<td>[কঁ] ] “ko” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>কু</td>
<td>[কু] ] “ku” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>কৃ</td>
<td>[কৃ] ] “kri” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>কো</td>
<td>[কো] ] “kou” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>কা</td>
<td>[কা] ] “ka” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>কি</td>
<td>[কি] ] “ki” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>কে</td>
<td>[কে] ] “ke” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>কঁ</td>
<td>[কঁ] ] “ko” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>কু</td>
<td>[কু] ] “ku” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>কৃ</td>
<td>[কৃ] ] “kri” (TEMP) &lt;.,0,0&gt;</td>
</tr>
<tr>
<td>কো</td>
<td>[কো] ] “kou” (TEMP) &lt;.,0,0&gt;</td>
</tr>
</tbody>
</table>

Similarly for all consonant it should need to entries in word dictionary.

Example 1-
Let’s an intermediate UNL expression-
```
{unl}
agt(read(icl>see>do,agt>person,obj>information).@entry.@present.@progress, "karim", TEMP :05)
{/unl}
```

Post converter firstly read the full sentence. When it finds the word “করিম” with attribute TEMP converter collect this word and push it into post converter. Then applying rules it convert into UNL word “karim”. Converter parses “করিম” letter by letter.

ক = ক + ঙ + r + i
arendra = r + i + m

That’s mean ‘করিম’ converted in “Karim”

Thus Post converter converts all proper nouns Bengali to English. Here we only try to present how post converter converts Bengali to English. In future we define the full for Bengali to English conversion.
6. RESULT ANALYSIS

To convert any Bangla sentence we have used the following files:
- Input file
- Rules file
- Dictionary

We have used an Encoder (EnCoL33.exe) and here I present some screen of en-conversion. Screen print shows the Encoder that produces Bangla to UNL expression or UNL to Bangla (Fig.6).

![Figure 6. Encoder for En-conversion](image)

When users click on convert button it generates corresponding UNL expression. The bellow screen shows this operation (Fig.7).

![Figure 7. En-conversion](image)

Based on the three steps pronoun detection technique we define rules for UNL system which identify the proper noun from a Bengali sentence and create relative UNL expression from the sentence.

6.1 Example:

This case is the form of a noun or pronoun used in the subject or predicate nominative. It denotes the agent of the action stated by the verb. For example, “কারিম পড়েছে” pronounce as Karim Portechhe means “Karim is reading”. Here subject Karim initiates an action. So, agent (agt) relation is made between subject “Karim” and verb “read”.

The following dictionary entries are needed for conversion the sentence:
- \{কারিম\} “karim (icl>name, io>person,com>male)(N)
- \{পড়ে\} “read (icl>see>do,agt>person,obj>information) (ROOT,CEND,^ALT)
- \{ইচেছে\} “” (VI, CEND, CEG1, PRS, PRG, 3P)

Where,

N denotes noun, ROOT for verb root, CEND for Consonant Ended Root, ^ALT for not alternative, VI is attribute for verbal inflexion, CEG1 for Consonant Ended Group 1, PRS for present tense, PRG for progress means continuous tense and 3P for third person.

But কারিম is the name of a person. So it is not needed to entry on dictionary. To convert this sentence into UNL expression when En-Converter finds word which is not in dictionaries then it creates a temporary entry. So if কারিম is not in dictionary then the temporary entry as like:

\{কারিম\} “karim “ (TEMP) <.0,0>: [ ]

**Applied rules:**

R[:::]|^PRON,^N,^VERB,^ROOT,^ADJ,^ADV,^ABY,^KBI V,^BIV:+N,+PROP::}(BLK)P10;

After applying these rules it is conceder as a proper noun. And temporary entry converted as:

\{কারিম\} “কারিম “ (N,PROP,TEMP) <.0,0>: [ ]

To convert this sentence into UNL expression morphological analysis is made between “পড়ে” (por) and “ইচেছে” (itechhe) and semantic analysis is made between “কারিম” (karim) and “পড়েছে” (poritechhe)

**Rule of morphological analysis:**

After applying some right shift rules when verb root “পড়ে” (por) comes in the LAW and verbal inflexion “ইচেছে” (itechhe) comes in the RAW then following rule is applied to perform morphological analysis:

+{ROOT,CEND,^ALT,^VERB:+VERB,-ROOT,+,@::} {VI, CEND::}P10;

**Rule of semantic analysis:**

After morphological analysis when “কারিম” appears in the LAW and verb “পড়েছে” (poritechhe) appears in the RAW the following agent (agt) relation is made between “কারিম”
(karim) and “পড়িঝেছে” (poritechhe) to complete the semantic analysis.

\[ \text{agt} \text{read(icl>see>do,agt>person, obj>information),@entry.@present.@progress, karim TEMP: 05) } \]

Where, N denotes the attributes for nouns or pronouns.

The intermediate language UNL form shows that the word “করিম” same as before conversion. When the sentence “করিম পড়িঝেছে” converted into another language then the word “করিম” is not understandable. Because of, in UNL expression the word “করিম” is Bengali word.

\[
\text{[unl]} \quad \text{agt}(\text{read(icl>see>do,agt>person, obj>information),@entry.@present.@progress, karim TEMP: 05)}
\]

So this expression is invoked into Post Converter which converts this word into English using proposed simple phonetic method.

Post Converter takes the UNL expression and converts the proper noun “করিম” Bengali to English “karim”.

\[
\begin{align*}
\text{ক} &= \text{ক} + ক \rightarrow \text{Ka} \\
i &= \rightarrow \text{ri} \\
\text{ম} &= \rightarrow \text{m}
\end{align*}
\]

That’s mean “করিম” converted in “karim”.

After conversion the final UNL expression is like as-

\[
\text{[unl]} \quad \text{agt(read(icl>see>do,agt>person, obj>information),@entry.@present.@progress, karim TEMP: 05)}
\]

Thus the UNL converter and Post converter can identify any proper noun from Bengali sentence and convert it corresponding UNL expression.

7. CONCLUSION

Here we have defined a procedure to identified proper noun from Bengali sentence and conversion method from bangle to UNL expression. We have also demonstrated how UNL converter identified proper noun from Bengali sentence and the UNL expression conversion by taking a sentence as an example. Here in result analysis section we demonstrate examples briefly. Here we define our work as two parts, firstly identified a proper noun from Bengali sentence and secondly convert this proper noun into UNL form. In the second parts we use a converter named as post converter which use simple phonetic method to convert Bengali to English and we only mention the simple procedure not complete. Our future plan in this regards is give the complete rules for post converter. We will also works on Bengali language and give the complete analysis rules for en-conversion program to convert Bangla to UNL language and generation rules for de-conversion program to convert any UNL documents to Bangla language.
REFERENCES

Configuration, Interfacing and Networking of Wireless IP-Based Camera for Real-Time Security Surveillance Systems Design

†V.A. Akpan
Department of Physics Electronics
Federal University of Technology
Akure, Ondo State, Nigeria.

R.A.O. Osakwe
Department of Physics
The Federal University of Petroleum Resources
Effurun, Delta State, Nigeria.

S.A. Ekong
Department of Physics and Energy Studies
Achiever’s University
Owo, Ondo State, Nigeria.

ABSTRACT

This paper presents three techniques for configuring, interfacing and networking of a wireless IP-based camera for real-time security surveillance systems design. The three different real-time implementation techniques proposed for configuring, interfacing and networking the IP camera are: 1). accessing the IP-based camera using the Wanscam vendor software, 2). accessing the IP-based camera via Firefox® web browser and 3). accessing the IP-based camera via MATLAB/Simulink®. The paper also demonstrates the interfacing and networking of the IP-based camera with multiple computers using an Ethernet switch. The live streaming video based on the proposed techniques can be adapted for image detection, recognition and tracking for real-time intelligent security surveillance systems design.

Keywords—Closed-circuit television (CCTV), Internet protocol (IP), security surveillance, wireless networking

I. INTRODUCTION

Traditionally, the surveillance monitoring system is done in the larger room and by a mount of manpower. But nowadays, monitoring surveillance system can be done through online network. This type of monitoring is less time consuming and can reduce the manpower. Moreover, it gives the user flexibility to monitor their properties wherever they want as long as they have the internet network. Security of lives and properties is an aspect of life which cannot be toyed with. Governments and individuals desire to know the conditions of their highly valued properties every second of life despite the fact that these properties are located in different places across the globe. Surveillance is the monitoring of the behaviour, activities, or other changing information, usually of people for the purpose of influencing, managing, directing, or protecting them [1]. Alternatively, surveillance can be the observation of individuals or groups by government organizations but can also relate to disease surveillance, which monitors the progress of a disease in a community while not directly observing individuals.

This work is concerned with the configuring, interfacing and networking of IP-based camera for real-time security surveillance systems design. The word surveillance may be applied to observation from a distance by means of electronic equipments such as CCTV cameras and IP camera, or interception of electronically transmitted information such as Internet traffic and/or phone calls. It may also refer to simple, relatively no- or low-technology methods such as human intelligence agents and postal interception. Surveillance is very useful to governments and law enforcement to maintain social control, recognize and monitor threats, and prevent/investigate criminal activity. With the advent of programs such as the Total Information Awareness program and ADVISE, technologies such as high speed surveillance computers and biometrics software, and laws such as the Communications Assistance For Law Enforcement Act, governments now possess an unprecedented ability to monitor the activities of their subjects [2].

†Corresponding Author: avincent@eng.auth.gr
There are different types of surveillance methods available in the world today, example of which are; Computer surveillance, telephones, social network analysis, biometric surveillance, aerial surveillance, data mining and profiling, corporate surveillance, human operatives, satellite imagery, identification and credentials, global positioning system, mobile phones and surveillance camera [3].

However, video surveillance systems have since the 1970s consisted of National Television System Committee (NTSC) or Phase Alternating Line (PAL) analogue cameras connected over a coaxial cable network to VHS tape recorders or digital video recorders (DVRs) in a monitoring station. Such surveillance systems are often comprised of black and white, poor quality analogue videos with little or no signal processing, recorded on the same cassette. Most of the recorded images are of insufficient quality to hold as evidence in a law court. It is also expensive to have human operators monitoring real-time camera footage all round the clock. The effectiveness and response of the operator is largely dependant on his/her vigilance rather than the technological capabilities of the surveillance system [4].

Events and activities can be missed, should the concentration level of the operator drop; attention levels drop significantly after 15 minutes of inactivity in the scene. The advent of high resolution digital IP surveillance cameras, connected via the internet to a remote security monitoring station, enables a new approach that draws attention to events identified in the camera scene. IP-based cameras coupled with the introduction of video content analysis or video analytics promise to extend the reach of video beyond security in a local area into a wide area surveillance system. Such automation and wider coverage will significantly reduce the drudgery workload on law enforcement agencies, thus making it possible for them to concentrate on the thing they do best: responding to suspicious events [5].

2. OVERVIEW AND DESCRIPTION OF THE WIRELESS WANS CAM IP-BASED CAMERA

2.1 Overview of IP-Based Camera: Characteristics and Specifications for Implementation

The surveillance operation by Internet Protocol (IP) camera is by streaming and live video streaming refers to sending video and audio signal captured by an IP camera in real time over the internet. Today, live video streaming technologies are widely used in broadcasting, connecting friends and relatives in online chat, online face to face business transaction, selling products and services, online teaching, showing online movie and finally streaming of live video is used in surveillance which is the purpose of this project. IP Cameras or network cameras were first introduced by Axis in 1996 [6]. They developed an open source programming language called VAPIX which utilized an embedded Linux operating system within the cameras. IP cameras transmit video using open internet protocols and standards (mainly HTTP) for recording or monitoring. The camera has a built in web server.

This means it can operate as a standalone product, accessed by its IP address and therefore does not require the host computer/workstation to function. This provides an economical method to monitor an area from a remote location. These cameras use the HTTP protocol to issue and receive instructions or data to or from an external network source, i.e. to view a Motion JPEG stream from the camera: http://192.168.1.126:81 (or a similar address) can be accessed through a standard http connection in code. The live video stream from the camera is obtained using the standard HTTP via the URLs with the following http://192.168.1.126:81. Various arguments can also be passed to the camera web server as long as the TCP/IP connection exists.

2.2 Network Bandwidth

IP Cameras typically generate between 0.2 Mbps and 2 Mbps of traffic on a network. This figure varies with size, frame rate and image compression used. A typical 100 Mbps office network can have a number of cameras attached, without effecting standard traffic. A larger number of cameras may require their own surveillance networks [7].

2.3 Camera Image Compression

IP cameras use either MJPEG (Motion Joint Photographic Experts Group) or MPEG 4 (Moving Pictures Expert Group) compression techniques for image transmission. MJPEG saves each frame as a JPEG whereas MPEG only retains information about the differences between the captured frames. JPEG is a JPEG based compression codec. It is the current top choice for video surveillance as it is inexpensive to implement and simple to decode, requiring a few modifications on top of existing JPEG compression systems. It does however require a large amount of bandwidth on a network, as much as a full T1 link for full motion video. MPEG 4 is primarily designed for video streaming from 4,800 bps to 4 Mbps and has higher quality compression than MJPEG. For the purposes of video surveillance, MPEG 4 would be the ideal choice, but it is costly to implement and can be very difficult to decode images from the stream. In this project, a Wanscam IP camera is used to capture the image.

2.4 The WanScam IP Camera

The Wanscam AJ Series IP Camera products are designed and equipped for local and remote network video surveillance system, including wired IP bullet camera, wireless IP bullet camera, IP IR dome camera, IP IR waterproof camera, IP Pan/Tilt/Zoom Camera etc. It adopts high performance chip to ensure high quality media processor which processes audio and video collection, compression and transmission. Standard M-JPEG compression format ensures clear and streaming video performance. It enables users to view live video via IE6.0, IE7.0, IE8.0, Firefox, Google browser or other standard browser.
2.5 The Wanscam AJ series Wireless IP-Based Camera’s Specifications

It adopts high Performance, strong function media processor 32-Bit RSIC (reduced instruction set computer); It has high sensor CMOS resolution; it adopts optimized MJPEG video compression algorithm, realize high-definition images transmission in narrow bandwidth; maximum support of up to 15 users viewing at the same time, no limit for users if using forwarder; several server functions; built in web server, convenient for users to use standard browse to realize the real-time monitoring and setting administration; support WIFI/802.11 b/g/n wireless networking; support remote system update; support DDNS analysis, support LAN & Internet (ADSL,Cable Modem); support variety of network protocol: TCP/IP, UDP, SMTP, PPPoE, dynamic DNS, DNS Client, SNTP, BOOTP, DHCP, FTP, SNMP, WIFI/802.11b/g; parts of modes products support for one/two way audio, talkback; support motion detection alarm function (area & sensitivity configurable); support image snapshot; Abnormal automatic recovery function, auto reconnection available when network; support for interruption occurred streaming video; dynamic alarm functions with alarm time-schedule reconfigurable facilities.

3. REAL-TIME IMPLEMENTATION OF THE IP-BASED WIRELESS CAMERA

3.1 Overview of the Real–Time Implementation of the IP-Based Camera

To access the IP camera for live video streaming or for real-time video transfer, either wirelessly or using hard wire connection, also, either for local area network connection (LAN) or for wide area network connection (WAN), there are ways through which this can be achieved. In this chapter, the software involved in interfacing the IP camera with the computer at the monitoring end and the stages of installation of each of the software till a live video is streamed or transferred are discussed. Below is the list of the ways to interface the IP camera: 1). using Wanscam vendor software, 2). using web browser, and 3). using MATLAB/Simulink® [8].

3.2 Accessing the IP-Based Camera Using the Wanscam Software

The software is installed through the Wanscam accompanied CD. When the CD is inserted into the host computer, the following options are itemized: Mobile view software, computer IE view via Ocx Setup.msi and the Search tool.exe. From the above options, Ocx Setup.msi is selected, clicked and run leading to the installation of the software. Thereafter, the camera powered before connecting the IP camera host computer via a CAT5E cable with the RJ45 connector at both tips of the cable. Then, the Search too.exe is double clicked and run, this will search for the IP of the connected IP camera and this page is loaded as shown in Fig. 1.
The software searches for the Wanscam camera IP address automatically over a LAN. In Fig. 1, the following actions take place: 1). Current Computer indicates the Computer’s IP Address information, 2). Camera information indicates the IP camera’s IP Address information, and 3). If found that the camera’s “Subnet Mask”, “Gateway”, “DNS Server” is not as same with the current computer’s. Then, there is a need to change either the camera’s IP address or the host computer IP address, because the IP addresses of the IP camera and the computer must be on the same network. It is important to ensure that the “Subnet Mask”, “Gateway”, “DNS Server” are the same as the router’s or as the current computer’s.

Configuration of the Network
Once the camera’s IP address’ Subnet Mask, Gateway, DNS Server are not the same as the computer or router, one needs to configure the camera’s Network parameters manually. The IP camera can be connected directly to the computer and it can at the same time be connected to the Router or switch and from any of this to the computer i.e. IP camera → Router/switch → Computer.

1). IP address: The IP address assigned is filled in and it is important to ensure that it is in the same subnet as the Gateway, and the subnet should be the same as the computer or router (i.e. the first three sections are the same).

2). Subnet Mask: The default subnet mask of the equipment is: 255.255.255.0. One can find the Subnet Mask from the host computer or router.

3). Gateway: It is important to ensure that it is in the same subnet with computer’s IP address. Here gateway is the LAN IP of the router.

4). Primary DNS: IP address of IPS network provider. One can also set it as the same as the Gateway.

5). Http Port: LAN port assigned for the equipment, default is 99. One can change the port number to any of this: 81, 98, 211, 9999, e.t.c.

Double clicking on the IP address of the camera in the Fig. 1 above will load a page displaying the live streaming video shown in Fig. 2.

3.3 Accessing the IP Camera via the Firefox® Web Browser

To access the camera by Firefox web browser directly, the camera’s IP address and the http number is typed in the address bar of the Firefox and enter key is pressed. In this study, the camera’s IP address is configured as 192.168.1.126:99 and the enter key is pressed, and the web page shown in Fig. 3 pops up.

After inputing the IP address in the dialogue box in Fig. 3, the enter key is pressed and the page in Fig. 4 pops up and the Username “admin” is entered while the Password is left blank and then ok is clicked. The resulting page shown in Fig. 5 pops up.

During the first time login, the camera will need ActiveX prompt, this is either done by downloading the OCX (or run in CD) and install first, then, Run Add-on is chosen, it is refreshed and login into ActiveX Mode( For Firefox Browser) is selected and it is Signed In and the page in Fig. 6 below pops up.

One of the main objectives of this work is to set aside all the aforementioned software used in accessing or interfacing the IP camera with the laptop except with MATLAB/Simulink®. And in achieving this, MATLAB as a powerful computer programming language is used to access and interface the IP camera with the host computer. The processes and the stages involved are explained below.
3.4 The MATLAB Implementation Platform

MATLAB (MATrix LABoratory) [8] is a numerical computing environment and fourth-generation programming language. Developed by The MathWorks, MATLAB/Simulink® allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages, including C, C++, Java, and Fortran. Although, MATLAB is intended primarily for numerical computing, an optional toolbox uses the MuPAD symbolic engine which allows access to symbolic computing capabilities. An additional package, Simulink, adds graphical multi-domain simulation and Model-Based Design for dynamic and embedded systems [9].

In 2004, MATLAB had around one million users across industry and academia. MATLAB users come from various backgrounds of engineering, science, and economics. MATLAB is widely used in academic and research institutions as well as industrial enterprises. It was first adopted by researchers and practitioners in control engineering, Little's specialty, but quickly spread to many other domains. It is now also used in education, in particular the teaching of linear algebra and numerical analysis, and is popular amongst scientists involved in image processing and this consequently makes it suitable for this project. Setting aside the software from the vendor in accessing the IP camera and using MATLAB to access it for live video either wired (LAN) or wirelessly (WAN) involves series of processes and stages.

Networking MATLAB with the IP Camera using Hard-Wired Connection

This is when the live video from IP camera is streamed directly to MATLAB interface on a laptop when they are at a short range of distance from each other. This can be achieved in two ways, namely:

1). Using MATLAB web browser (imread method)

In this method, the IP camera is registered in the image acquisition tool box of MATLAB by creating an adaptor kit for the IP camera.

2). Using MATLAB web browser (imread method)

In this second method, the IP addresses of the IP camera and the host computer are ensured to be on the same network. If, however, this is done, then MATLAB will be made to run on the laptop that will receive the live video.
3.5 Accessing the IP-Based Camera Over a Network With Multiple Computers Using 100MB/s Ethernet Switch

Using any of the aforementioned software, either the software from the vendor, web browser or the MATLAB, the software is installed first in each of the computers with the same procedure as other ways explained above. Each of the computers is connected to the switch using CAT5E cables with RJ45 connector at the tips of the cables. The switch is then connected to IP camera using the same cable and connector, as shown in the Fig. 12. Each of the computers can access the camera either simultaneously or separately. The IP addresses of the involving computers must be on the same network for to communicate over LAN.

4. CONCLUSION

Real-time capturing and viewing streaming video from an IP-based camera on the host computer has been proposed and implemented using three different methods, namely: 1) using the software from the vendor, 2) using the web browser, and 3) using MATLAB/Simulink.

As a future direction, work has started on techniques to capture streaming video using MATLAB/Simulink® image acquisition and signal processing toolboxes with a view for the design and implementation of an online detection and recognition of multiple targeted images in real time for intelligent security surveillance systems design and deployment.
REFERENCES

Author’s Brief
Vincent A. Akpan (M’08) obtained the B.Sc. degree in Physics from the Delta State University (DELSU), Abraka, Nigeria in 1997; a Master of Technology (M.Tech.) degree in Instrumentation from The Federal University of Technology, Akure (FUTA), Nigeria in 2003; and a Ph.D. degree in Electrical & Computer Engineering from the Aristotle University of Thessaloniki (AUTH), Thessaloniki, Greece in 2011. Between 1998 and 2004, he was a Graduate Assistant with the Department of Physics, DELSU. Since 2004 he has been with the Department of Physics Electronics, FUTA where he is currently Lecturer I. His integrated research interests include: computational intelligence, nonlinear system identification, adaptive predictive control, real-time embedded systems, signal processing & machine vision, instrumentation, and Robotics & Automation. He is the co-author of a book and has authored or co-authored over 40 articles in refereed journals and conference proceedings. He is a regular reviewer in several local and international scientific/academic journals and several IEEE sponsored conferences. He is the Managing Editor of Nigeria Journal of Pure and Applied Physics as well as an editorial board member for the American Journal of Intelligent Systems. Dr. Akpan is a member of the IEEE, USA, The IET, UK and a Fellow of IPMD, Nigeria. He was one among two Nigerian recipients of the 2005/2006 Greek State Scholarship (IKY) for a Ph.D. programme tenable in Greece.

Reginald A. O. Osakwe received his Ph.D in electronics from the University of Ibadan, Nigeria in 1995. Since 2007 he has been with the Department of Physics, The Federal University of Petroleum Resources, Effurun, Delta State. Nigeria where he is currently a Senior Lecturer. During the periods 1998 – 2009 he has been a visiting lecturer to the Shell Intensive Training Programme (SITP), Warri, Delta State under the Shell Petroleum Development Company (SPDC) of Nigeria. His research interests are on microprocessor architecture, embedded computer systems, chaotic dynamics, and image and signal processing systems. Dr. Osakwe is a member of the Nigeria Institute of Physics (NIP). He was a recipient of The Federal Government of Nigeria Scholarship to study in the U.S.A. leading to the award of an M.Sc.
Sylvester A. Ekong had his B. Sc. Degree in Physics (With Electronics) from the Delta State University (DELSU), Abraka, Nigeria in 2005; and a Master of Technology (M. Tech.) degree in Condensed Matter Physics from The Federal University of Technology, Akure (FUTA), Nigeria in 2011. Between 2006 and 2007 he was a Physics and Further Mathematics tutor at Government Model Secondary School, Ajikamai, Shendam, Plateau State, Nigeria; Between 2007 and 2009 he was a tutor of Physics and Further Mathematics and also the practical Physics coordinator at St. John’s College, Kachia, Kaduna State, Nigeria; He was, between 2009 and 2011, a tutor and practical Physics coordinator with Idris Primier College, Akure, Ondo State; and also between 2011 and 2012 he was a Physics and Further Mathematics tutor with Model secondary school, Alagbaka, Akure, Ondo State, Nigeria. After obtaining his M. Tech. degree, in November 2011 he joined as an Assistant lecturer in 2012, the Department of Physics and Energy Studies, Achievers University, Owo, Nigeria. He has co-authored over 15 articles in refereed journal and conference proceedings. His research interests are on — digital signal processing, soft-computing, Theoretical & Computational condensed matter and chaotic systems. Mr. Ekong is a member of The African Review of Physics (ARP) and The National Association of Theoretical and Mathematical Physics (NAMP). He is currently a Ph.D. student in the Department of Physics, University of Benin, Benin City, Nigeria.
An Investigation into the Conceptual Controversies between Artificial Intelligence and Computational Intelligence

K.T. Igulu
Department of Computer Science
Rivers State Polytechnic, Bori
Rivers State, Nigeria.
E-mail: igulukt@gmail.com,
Phone: +2347018827522

Z.P. Piah
Department of Computer Science
Rivers State Polytechnic, Bori
Rivers State, Nigeria.

P.O Asagba PhD
Department of Computer Science
University of Port Harcourt
Rivers State, Nigeria.

ABSTRACT

Artificial intelligence (AI) is one of the oldest and best known research fields in computer science which is aimed at giving intelligence in machines. In spite of enormous effort geared towards AI, its boundary and interference to other fields are yet undefined. The controversy in AI’s boundary has bred some concepts which are in no practical sense different from the goal of AI. Most of these concepts are: computational intelligence (CI), soft computing (SC), natural computing (NC), machine learning (ML), approximating reasoning (AR), fuzzy logic (FL) and adaptive computing (AC). There is still not an objective boundary of these concepts. This paper investigates the conceptual controversies between AI and CI. AI and CI are major contending concepts that have posed confusions to researchers and students. We have cited some definitions and comments about AI and CI from trusted sources. After thorough analysis of the definitions and comments, our own view of what AI and CI are is presented. This paper does not give extensive information about AI and CI approaches, tools and applications. Our emphasis is in distinguishing AI from CI. An extensive bibliography is given for references.

Keywords: artificial intelligence; computational intelligence; soft computing; nature-inspired; controversies.

I. INTRODUCTION

What exactly are these terms Artificial intelligence (AI) and computational intelligence (CI) and how are they the most fascinating research areas in the computing discipline? To provide qualitative answer to this question, it is consequential to highlight the concept of intelligence. Although intelligence is not the goal of this paper but any paper about AI and CI cannot do away with the concept of intelligence. The main controversy in giving a precise definition to AI and CI is the inclusion of intelligence. The study of (human) intelligence has a rich history over three millenniums [4]. In spite of the ages it has undergone, the definition of intelligence is still debatable. Intelligence has been defined in different ways, including the abilities, but not limited to, abstract thought, understanding, being self-aware, communication, reasoning, learning, having emotional knowledge, retaining, planning, and problem solving[1]. Another school of thought segments intelligence into:

a. Rational thought and reasoning-The ability to act purposefully in an environment.
b. The ability to deal with situations, in an effective manner, within an environment.
c. Cognitive – Examples of cognitive ability: memory, perception, concept formation, problem solving, mental imagery, action, association, language and attention.
d. The ability to learning from experience.
e. The ability to live and cope with the demands of daily life [20].

In this paper, we will also adapt the rational agent approach of intelligence in [19].
AI is primarily the science of creating intelligent systems one of the largest scientific communities. AI has been established already in circa 1950, working on problems that require intelligence to be solved. It's evolution has been summarized in the 25th anniversary issue of the *AI Magazine* by Mackworth[2]: “In AI’s youth, we worked hard to establish our paradigm by vigorously attacking and excluding apparent pretenders to the throne of intelligence, pretenders such as pattern recognition, behaviourism, neural networks, and even probability theory. Now that we are established, such ideological purity is no longer a concern. We are more catholic, focusing on problems, not on hammers. Given that we do have a comprehensive toolbox; issues of architecture and integration emerge as central [5].”

IEEE Computational Intelligence Society defines CI as consisting of neural computing (NC), fuzzy computing (FC) and evolutionary computation (EC). The approach taken by researchers and book authors is to treat computational intelligence as an umbrella under which more and more methods will be added. A good definition of the field is therefore impossible, because different people include or exclude different methods under the same CI heading [5].

2. ARTIFICIAL INTELLIGENCE (AI)

2.1 Brief History of AI

In spite of the wide usage and applications, journal papers and books written about AI, there is still not simple and all-encompassing meaning of AI. Before exploring into the concept of AI, we will highlight its origin and intercepting fields of AI. As fuzzy as its definition so is the origin of AI. Some scholars attribute the origin of AI to Aristotle for his works on logic and deductive reasoning (syllogism). Though his works on logic has great impact in AI but deductive reasoning alone does not imply intelligence. Another School of thought attributes the origin of AI to Alan Turing for his work on Turing machine. Turing defined intelligent behavior as the ability to achieve human-level performance in all cognitive tasks, sufficient to fool an interrogator.

Roughly speaking, the test he proposed is that the computer should be interrogated by a human via a teletype, and passes the test if the interrogator cannot tell if there is a computer or a human at the other end [19]. Although formidable, but has not got necessary qualities of intelligence as defined above. His work was a major motivation to researchers. His work could be viewed as machine that acts humanly [19]. Most AI literatures ignore the influence of Von John Neumann works of automata on AI. One of his noted comments is that “Natural organisms are, as a rule, much more complicated and subtle and therefore much less well understood in detail, than are artificial automata. Nevertheless, some regularities which we observe in the organization of the former may be quite instructive in our thinking and planning of the latter; and conversely a good deal of our experiences and difficulties with our artificial automata can be to some extent projected on our interpretations of natural organism ” [5]. He envisaged what is seen today as Soft computing (SC) a term used interchangeably as CI.

With much effort invested than noticed, a new wave of AI came into picture at a conference on the campus of Dartmouth College in the summer of 1956. The attendees, including John McCarthy, Marvin Minsky, Allen Newell and Herbert Simon, became the leaders of AI research for many decades. Parallel to what is known as rule-based AI was the advancement of the nature-inspired techniques grouped under the name CI.

A belief in the efficacy of stepping through simple rules, combined with a belief that the operations of the Turing machine accurately model human behaviour, led to claims for behaviour achievable in digital computers which could not be substantiated. The turning point in this phase seems to have been the project for automatic translation of languages. This failed so badly that it provoked a complete cessation of funding for AI research in Great Britain and a severe reduction in the United States. The unsatisfactory outcome of projects undertaken in the 1950’s caused the whole field of AI research to become relatively dormant until the mid-1960. Soon after its renascence, impressive machine behaviour was achieved, especially in problem solving and diagnostic assistance. The ability to reflect the behaviour of an expert was impressive, even though the range of behaviour was quite limited. Landmarks of this period include:

a. development by McCarthy [9] of the LList Processing (Lisp) language, which focuses on recursive processing of arbitrary lists;

b. the “blocks world,” from Understanding Natural Language, by Terry Winograd [21], which demonstrates limited communication with a robot in natural English and with automated planning;

c. The Eliza program, which mimics a Rogerian psychotherapist, by Weizenbaum [10];

d. The Semantic Organization of Memory, by Quillian [14], which outlines a system of world linkages reflecting concept linkages in natural language;

e. The MYCIN program [7], which demonstrated the utility of rule sets and confidence factors in developing a diagnostic adviser for physicians.

The subject of AI has evolved, and continues to evolve, from the intersection of research from such fields as cognitive science, philosophy, psychology, mathematics, linguistics, statistics, physics, biology, engineering, ontology, finance, economics and natural sciences. AI is forecasted to have applications and interferences to almost all fields of study in few decades.
2.2 Cited Definitions and Commentaries of AI

The phrase AI, which was coined by John McCarthy about five decades ago, sidesteps a concise and formal definition to date. This is ephemeral. One representative definition is pivoted around the comparison of intelligence of computing machines with human beings. Another definition is concerned with the performance of machines which "historically have been judged to lie within the domain of intelligence". None of these definitions or the like has been universally accepted, perhaps because of their references to the word “intelligence”, which at present is an abstract and immeasurable quantity. A better definition of AI, therefore, calls for formalization of the term “intelligence”. Psychologist and Cognitive theorists are of the opinion that intelligence helps in identifying the right piece of knowledge at the appropriate instances of decision making. The phrase “AI” thus can be defined as the simulation of human intelligence on a machine, so as to make the machine efficient to identify and use the right piece of “Knowledge” at a given step of solving a problem. A system capable of planning and executing the right task at the right time is generally called rational. Thus, AI alternatively may be stated as a subject dealing with computational models that can think and act rationally[19][3].

Another paradigm in the definition of AI was proposed in [19] which seem to be simplistic but much more complicated than expected. Here AI is segmented into four intersecting categories:

a. thinking humanly
b. acting humanly
c. thinking rationally and
d. Acting rationally.

The following definitions were the motivation for the categorization in[19]: "The exciting new effort to make computers think . . . machines with minds, in the full and literal sense", "[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning ...", "The art of creating machines that perform functions that require intelligence when performed by people, "The study of how to make computers do things at which, at the moment, people are better", "The study of mental faculties through the use of computational models", "The study of the computations that make it possible to perceive, reason, and act", "A field of study that seeks to explain and emulate intelligent behaviour in terms of computational processes", "The branch of computer science that is concerned with the automation of intelligent behaviour". The two approaches emphasized in [19] is human-centred and rationalist approaches. Much emphasis was laid on the latter. A system is rational if it does the right thing [19]. The right thing is usually the one which best (or at least well) helps the system in achieving its goals[19]. In [19], intelligence or rationality is based on subjective and dynamic reasoning rather than the ability to correctly solve problems irrespective of complexity. The definitions posited by McCarthy and IEEE seem to be the basis from which all other views originate but this is also ephemeral. Although not yet pragmatic, the actualization of strong AI will contradict the IEEE, 1996 view of AI. The future of AI is not deterministic. It is quite difficult to extend this definition, because the definition of what factors describe human intelligence is not clear.

Two types of methodology have become dominant in AI. The first is concerned only with sets of rules and their activation. The second is concerned with more general problems of the organization of knowledge, plus a need for cooperation between various sources of knowledge in a synergistic way. Systems using only the rule-based approach focus on capture of appropriate rules, the facts used, and proper rule execution in a limited, specific situation. These are the kinds of systems properly referred to as expert systems, since they focus on capturing the insights of an expert in a small domain, and coding these insights. The second type of system encompasses a larger domain than just facts and rules for one situation. The latter type of system is better referred to as a knowledge-based system.

3. COMPUTATIONAL INTELLIGENCE(CI)

Parallel to what is known as rule-based AI was the advancement of the nature-inspired techniques grouped under the name CI. Although used fairly widespread [3], there is no (yet) on what CI exactly is, but there is a widely accepted view on which areas belong to it: EC, FC and NC.

The World Congress on CI held every four years ([1994 Orlando, 1998 Anchorage, and 2002 Honolulu]) consists of three tracks, the IEEE International Conference on EC, FC, and NC [4].

From 1956 to 1969 much research was done in modeling biological neurons. Most notable were the work on perceptrons by Rosenblatt, and the adaline by Widrow and Hoff. In 1969, Minsky and Papert caused a major setback to NC research. With their book, called Perceptrons, they concluded that, in their "intuitive judgment", the extension of simple perceptrons to multilayer perceptrons "is sterile". This caused research in NC to go into hibernation until the mid-1980s. During this period of hibernation a few researchers, most notably Grossberg, Carpenter, Amari, Kohonen and Fukushima, continued their research efforts.

© 2015 Afr J Comp & ICT – All Rights Reserved - ISSN 2006-1781
www.ajocict.net
The resurrection of NC research came with landmark publications from Hopfield, Hinton, and Rumelhart and McLelland in the early and mid-1980s. From the late 1980s research in NC started to explode, and is today one of the largest research areas in Computer Science. The development of evolutionary computing (EC) started with genetic algorithms in the 1950s with the work of Fraser. However, it is John Holland who is generally viewed as the father of EC, most specifically of genetic algorithms. In these works, elements of Darwin’s theory of evolution were modeled algorithmically. In the 1960s, Rechenberg developed evolutionary strategies (ES). Research in EC was not a stormy path as was the case for NC.

In 1920 Lukasiewicz published the first deviation from two-valued logic in his work on three-valued logic - later expanded to an arbitrary number of values. The quantum philosopher Max Black was the first to introduce quasi-fuzzy Sets wherein degrees of membership to sets were assigned to elements. It was Lotfi Zadeh who contributed most to the field of fuzzy logic, being the developer of fuzzy sets in 1965[12][13].

A key player in the field of CI is J.C. Bezdek who was the first researcher to publish a paper on the definition of CI. In one of his paper, he posited that: “… (strictly) computational systems depend on numerical data supplied by the manufactured sensors and do not rely upon “knowledge”.” Later in 1994, Bezdek offers CI as a” low-level computation in the style of the mind” and AI as “mid-level computation in the style of the mind”. He outlined that mid-level system include knowledge whereas low-level does not [11][4]. This school of thought implies that CI systems employ sensor data and the AI should be reserved for architectures that have a clearly identifiable non-numerical component or knowledge. Here CI systems are identified by the fact that: “It deals only with numerical (low-level) data, has a pattern recognition component, and does not use knowledge in the AI sense; and additionally, when it (begins to) exhibit (i) computational adaptively; (ii) computational fault tolerance; (iii) speed approaching human-like turnaround, and (iv) error rates that approximate human performance.” [4].

Another notion related to Bazdek philosophy but differs with the introduction of adaptation views CI as thus; “...Computational intelligence is defined as a methodology involving computing (whether with a computer, wetware, etc.) that exhibits an ability to learn and/or deal with new situations such that the system is perceived to possess one or more attributes of reason, such as generalization, discovery, association and abstraction. The output of a computationally intelligent often includes predictions and/or decisions…”

As cited in [4], one common denominator in the definitions of CI is adaptation. [4] puts it in perspective as “Any system... that generates adaptive behavior to meet goals in a range of environments can be said to be intelligent. In contrast, any system that cannot generate adaptive behavior and can only perform in a single limited environment demonstrates no intelligence.” It exceeds the scope of this paper to go into the investigation of all the aspects of CI such as NC, FC and EC. Extensive references have been provided to aid readers.

3.1 Computational Intelligence versus Soft Computing

Contrast to traditional hard computing, SC exploits the tolerance for imprecision, uncertainty, and partial truth to achieve tractability [2], robustness, low solution-cost and better rapport with reality”. SC as coined by Zadeh [13] [14] was intended to differentiate all techniques that mimic the human mind from the crisp computing. There is no difference in goal between CI and SC both is umbrella concept for nature-inspired problem solving techniques. Nevertheless, the slight difference between CI and SC is the composition of techniques. CI is a composition of EC, NC and FC whereas SC extends to some other techniques to probabilistic computing and rough set. In fact there is no boundary for SC techniques. CI is a subset of SC. Therefore we do not interchangeably use SC and CI in this paper. Figure 1 depicts this distinction.

4. COMPUTATIONAL INTELLIGENCE VERSUS ARTIFICIAL INTELLIGENCE

This section is the major motivation of this paper. The relationship between CI and AI has formed a frequently discussed issue during the development of CI. The huge majority of CI/AI researchers concerned with the subject sees them as different areas, where either [3]

- CI forms an alternative to AI
- AI subsumes CI
- CI subsumes AI

Supporting information to the first point could be found in Mark [1993[4]: “Although seeking similar goals, CI has emerged as a sovereign field whose research community is virtually distinct from AI”. Another school of thought that lays emphasis on point 2 above is the publication of Bazdek of 1994. He depicted CI to be a subset of AI. Bezdek summarized the relationships among components of intelligent systems with a figure, after Bezdek’s elaboration on his first definition in 1994. He describes three levels of system complexity, level A, B and C. Level A stands for artificial or symbolic, level B for biological or organic, and level C stands for computational or numeric systems [4]. He confirms his view positing that NC, FC and EC are AI enabling techniques.
Another notion that contradicts Bezdek’s view is proposed by Fogel in 1995 who posited that the basis of this distinction between AI and CI is the identification of adaptation as the key feature of intelligence. His argument could be summarized in the comments in [4]. Fogel[6] observes that the central focus in traditional AI research has been on emulating human behavior by extracting rules and knowledge from human experts. Furthermore, the vast majority of AI programs has nothing to do with learning. Traditional symbolic AI systems do not adapt to new problems in new ways, therefore they emphasize “artificial” and not the “intelligence”. They may play excellent chess, but in essence they are but complicated calculators. In contrast, CI techniques model natural processes or end-products associated with intelligent behaviors, either at the level of neuronal activity and function, human behavior, or evolutionary learning in the terms of adaptive behavior or adaptive genetics. Pushing it to the extreme, from these premises it may be implied that (traditional) AI systems are not intelligent, while CI systems are.

5. DISCUSSION

Figure 1. shows the categorization of the concepts. The AI is primarily a discipline oriented to creating intelligent systems. By this, any discipline that deals also with machine intelligence is subsumed by AI. We view the traditional AI approaches or techniques as deductive. Precisely, they follow the famous bottom-up approach. Techniques under this category cannot efficiently deal with acute and nondeterministic problems. They cannot deal with uncertainty and vagueness. Another categorization is the nature-inspired techniques. These techniques are believed to deal with the limitations of the counterpart category naturally. SC is seen as a super set which contains all techniques categorized as CI and more.

6. CONCLUDING REMARKS/FUTURE RESEARCH

The CI discipline is apparently at its infancy with respect to practical application of its constituent technologies. Any notion given at this stage will have to pass through a test of time. Below are our views of what CI should be and its relationship to other concepts.

- **AI:** In our view, **AI is the umbrella discipline of all disciplines that have to do with simulation of intelligence of any degree in machines.**
- **CI:** We define CI as an umbrella name for natured-inspired problem-solving techniques that could be used to simulate intelligence into machine or make a machine more intelligent or could be used in solving nonalgorithmizable problems. **In our view, CI is a discipline of its own whose constituent problem-solving techniques could be utilised to achieve a more efficient, error-tolerance, robust and economic AI entities.** Having techniques that can handle uncertainty, partial truths, ambiguities and imprecisions that model the nonlinearity and nondeterminism of the natural world, CI is the need of AI. CI techniques are inductive in nature(bottom-top) whereas the traditional AI techniques are deductive(top-down). Traditional AI also known as the rule-based AI employs crisp techniques and thus will pose difficulties in dealing with acute problems that require flexibility. These points imply that CI majors in solving complex or non-algorithmizable problems.

- We can vividly differentiate CI and AI from the view point that CI is an independent umbrella of techniques that can be used to realize AI systems whose intelligence could compete or/and outrun human intelligence. This also gives viability to the concept of strong AI. We should note that just as rule of thumb, heuristics and if...then approaches are techniques to achieve intelligence in a deductive manner so are CI techniques which employ the principle of inductive procedure.

As future research, investigation of the CI and AI constituent techniques with respect to practical works should be made to give direction to researchers and students delving into this area of machine intelligence. Detailed definition of terms and discussion on these concepts from an application perspective should also be included in such investigation.

Figure 1. Categorization Of The Concepts

Acknowledgment

We acknowledge the researchers whose works were cited in this piece.
REFERENCES

[10] J. Weizenbaum, Computer Power and Human Reason. San Francisco: W. H. Freeman & Co., 1976 (the inside story from the creator of the Eliza program cited in the literature, with cogent comments on some of the “reasoning” programs now available, such as the Macsyma program of Symbolics).
A Real-Time Voice-Control Model to Access Protected Resources

W.O. Adesanya
General Studies Department
Federal College of Agriculture
Akure, Nigeria.
Adjunct Lecturer at the Department of Computer Science
Tai Solarin University of Education
Ijagun, Ijebu-Ode. Nigeria. sanyalanre2003@gmail.com

O. B. Alaba & G. O. Odulaja
Department of Computer Science
Tai Solarin University of Education
Ijagun, Ijebu-Ode. Nigeria
alabaob@tasued.edu.ng; goddyseyi@gmail.com

ABSTRACT

The paper presents a model developed for real-time speaker recognition system that can be used to access restricted information or resources using human voice. Speaker recognition involves identification and verification of the speaker. At each stage, the voiceprint is compared with model voices of all speakers in the database. The comparison is a measure of the similarity (score) from which rejection or acceptance of the verified speaker is chosen. The speaker recognition noise cleaning techniques used were the Linear Predictive Coding (LPC) and Cepstral analysis for feature extraction. The system was trained and tested using a population of ten users, with additional ten impostors. The average call success of a true speaker was 96.5%. The impostor’s success rate was found to be 0%. The impostor’s success rate was statistically computed to be 0.0036%, which shows that the speaker recognition system exhibited good performance in real-time operation.

Keywords: Speaker Recognition, Real-Time, Linear Predictive Coding, Cepstral Analysis

I. INTRODUCTION

Naturally, spoken language is used by humans to communicate information. The voice signal conveys linguistic and speaker information plus the information about the environment in which the voice was produced and transmitted. Speaker recognition is a biometric modality that uses an individual’s voice for recognition purposes. The speaker recognition process relies on features influenced by both the physical structure of an individual’s vocal tract and the behavioural characteristics of the individual. Handwritten signature based verification has long been a common method within modern society, and more than ever Personal Identification Number (PIN) codes are being used for accessibility to homes, vehicles and ATM service points and the like. The rapid rise of technology has enabled the pursuit of even more complex biometric identity verification methods. Nowadays, the fields of science and technology pursue identification methods using individual biological characteristics such as DNA analysis, fingerprints, and voice analysis [17].

Speaker Recognition system can make it possible to use a person’s voice to control the access to restricted services (automatic banking services), information (telephone access to financial transactions), or area (government or research facilities). It also allows detection of speakers, for example, voice-based information retrieval, recognition of perpetrator on a telephone tap, and detection of a speaker in a multiparty dialog. It is a known fact that accessing protected resources is always carried out through the use of personal tokens like a key or badge, knowledge of certain information like a password or combination of numbers [1]. It is however observed that these tokens can be lost, stolen or counterfeited, thereby posing a threat to security. Thus, in order to reduce this security threat, this paper focuses on real-time voice-control strategy to access the restricted resources, since voice is unique to each person and cannot be lost or stolen. Moreover, voice-driven based solutions are found to be able to provide for confidential financial transactions and personal data privacy, even though the result of the survey conducted by the International Biometric Group (IBG) in 2004 confirmed that Voice identification is not widely used technique of biometric identification in the world today [3].
The remaining section is organized as follows: section 2 reviews a number of relevant literatures on speaker recognition system; section 3 describes the methodology for the proposed system while 4 and 5 describe the results and concludes the work respectively.

2. RELATED WORKS

There are several research works in the application of techniques and models used in extracting voice feature or matching feature in order to identify and verify speakers in Speaker Recognition System (SRS). A number of such relevant researches were reviewed in this paper. In [9], verification system authenticates a person’s identity by comparing the captured biometric characteristic with its own biometric template(s) pre-stored in the system which conducts one-to-one comparison to determine whether the identity claimed by the individual was true. A verification system either rejects or accepts the submitted claim of identity and that the identification system recognizes an individual by searching the entire template database for a match which conducts one-to-many comparisons to establish the identity of the individual. The delimitations of [9] were that the rate of fingerprint capture and feature extraction were not considered, although in a real-time world scenario, this is an important factor.

In [12], a stochastic model was developed to solve the problem of speech processing in speaker recognition. The research was able to develop a high-quality, multivariate and Hidden Markov Model (HMM) by means of Hidden Markov Toolkit (HMTK) tool software to determine the speaker but provision for grammar testing enlargement as the new models are needed for the new words training. However, the limitations of the research were the direct counting of the probability was very complicated; and that the current state depends on the previous state. A new feature selection method for speaker recognition was proposed by [11] to keep the high quality speech frames for speaker modelling and to remove noisy and corrupted speech frames.

The research adopted spectral subtraction algorithm to estimate the frame power. An energy based frame selection algorithm was then applied to indicate the speech activity at the frame level. The research was able to use the eigenchannel based GMM-UBM speaker recognition system to evaluate the proposed method. However, the research required long-term spectral analysis and computation found to be complex. [20] concentrated on optimized speech processing in the DSP56001 hardware platform, especially in the application of noise reduction and speech enhancement. [14] worked on a hardware based speech recognition system. Both work by [14, 20] were hardware based but were not concentrated in the area of speaker recognition, which is the focus of this paper, based on the observation that the size of the speaker database grows when the number of speakers in a system is increased.

3. METHODOLOGY

A voice-driven system involves two phases. In the first phase, a user enrols by providing voice samples to the system. The system extracts speaker-specific information from the voice samples to build a voice model of the enrolling speaker, Figure 1.

![Figure 1: Model of a Voice Identification System](image1)

In the second phase, a user provides a voice sample (also referred to as test sample) that is used by the system to measure the similarity of the user’s voice to the model(s) of the previously enrolled user(s) and, subsequently, to make a decision. In a speaker identification task, the system measures the similarity of the test sample to all stored voice models. In speaker verification task, Figure 2, the similarity is measured only to the model of the claimed identity.

![Figure 2: Model of a Voice Verification System](image2)
Several conversational telephone calls in English and Yoruba languages were conducted and recorded. The collected voices were processed through the use of notebook computer with an external microphone attached, where all the voices were recorded digitally into the computer via the microphone. Voice sampling was required to convert an analogue signal into a discrete signal, to be digitally processed by a digital computer. Further pre-processing such as speech framing, edge detection and windowing were performed to improve the raw digitized signal to be used in the feature extraction process, further steps taken as shown in Figure 3.

A digital signal processor running at 50MHz was used to execute the voice recognition algorithm. The Linear Predictive Coding (LPC) Cepstral technique was used for feature extraction of speech signal as the speech sample \( s(t) \) at time \( t \) was approximated as a linear combination of the past \( p \) samples

\[
a(t) = a_1 s(t-1) + a_2 s(t-2) + \ldots + a_p s(t-p) \tag{1}
\]

where the coefficients \( a_1, a_2, \ldots, a_p \) were assumed constant over a single speech frame. The autocorrelation method with function

\[
\rho[m] = \sum_{i=1}^{T} s(i) s(i+m), \quad m = 0, 1, 2, \ldots, p \tag{2}
\]

was used for estimating the coefficients which provided the energy of the speech frame, and was used for discarding silent frames. The LPC coefficients \( a_i(t) \), \( 0 \leq i \leq T-1 \) were computed from the autocorrelation vector using a recursion method known as Durbin’s method where the equations were solved recursively for \( i = 1, 2, \ldots, p \). On completion of the algorithm, the final solution was given as:

\[
a_{\text{LPC}} = \mathbf{LPC \ coefficients} = a_{ij}, \quad 1 \leq m \leq p \tag{3}
\]

Vector Quantization (VQ) codebook was used for feature matching, to efficiently represent speaker specific characteristics. One codebook was created for each \( i \) speaker during the training stage. During recognition, the total distance for the \( i \)-th speaker was computed by:

\[
D_i^j = \sum_{i=1}^{N_c} \sum_{j=1}^{C_i} d(y_i, C_j) \tag{4}
\]

where \( C_j \) is the \( j \)-th code vector of the \( i \)-th speaker’s codebook, \( N_c \) is the codebook size, \( y_i, y_2, \ldots, y_T \) represent the feature vector of the test utterance, \( D_i \) is the matching score and \( d(y, C_j) \) the distance between the feature vector and the codebook vector, where the speaker identification decision was based on the matching score. The speaker model with the smallest matching score, \( D_i \) was accepted as the producer of the voice sample, otherwise, rejected.

Speaker identification using Dynamic Time Warping (DTW) was implemented using a training or reference template for each speaker. During identification stage, a DTW score of the test utterance was made against each training template. Speaker identification was carried in favour of the speaker whose training template produced the lowest score, provided the score is within the threshold value. For speaker verification application, the test utterance was compared against the training template of the speaker who was being verified. The obtained DTW score was compared against a threshold value and the user was only verified if the score was lower than the threshold value set for the speaker. The verification threshold \( T \) was computed using this equation:

\[
T = H_{\text{unp}} + \mu_{\text{pk}} \sigma_{\text{pk}} = H_{\text{unp}} \sigma_{\text{pk}}^2 + \mu_{\text{unp}} \tag{5}
\]

Where (i) The mean, \( \mu_{\text{pk}} \) and standard deviation, \( \sigma_{\text{pk}} \) is computed from the DTW score from each digit; and

(ii) The mean, \( H_{\text{unp}} \) and standard deviation, \( \sigma_{\text{unp}} \) is computed from the DTW score against this user’s template and speech samples of impostors.

![Figure 3: Block diagram of software modules developed for the voice recognition system](image-url)
For the real-time experiment, the memory required for the types of classifier implementations were noted along with the execution time. The execution time was only given for the classifier training and recognition routine. The memory and processing time results were recorded. Voice access was only granted if both identification and verification were successful. An application software was developed using C programming language with Code Composer Studio (CCS) to generate the source codes for autocorrelation analysis, LPC Cepstrum and DTW. The DSP Starter Kit (DSK) Debugger was used to download source code to the speaker recognition system, which executes decoding and monitoring.

The average identification success rate and average verification success rate for both original speakers and impostors were given in percentage. The system performance was evaluated using Equal Error Rate.

4. DISCUSSION OF RESULTS

The speaker identification and verification result of a true speaker is given in Table 1. The average identification success rate was 96%, and average verification success rate was 97%. The overall call success result of a true speaker is given in Table 2. The average call success rate for a true speaker was 96.5%.

Table 1: The Voice identification and verification success count

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Successful Identification (True Acceptance)</th>
<th>Unsuccessful Identification (False Rejection)</th>
<th>Successful Verification (True Acceptance)</th>
<th>Unsuccessful Verification (False Rejection)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S 1</td>
<td>47</td>
<td>3</td>
<td>47</td>
<td>3</td>
</tr>
<tr>
<td>S 2</td>
<td>50</td>
<td>0</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>S 3</td>
<td>45</td>
<td>5</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>S 4</td>
<td>50</td>
<td>0</td>
<td>49</td>
<td>1</td>
</tr>
<tr>
<td>S 5</td>
<td>47</td>
<td>3</td>
<td>48</td>
<td>2</td>
</tr>
<tr>
<td>S 6</td>
<td>47</td>
<td>3</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>S 7</td>
<td>49</td>
<td>1</td>
<td>44</td>
<td>6</td>
</tr>
<tr>
<td>S 8</td>
<td>48</td>
<td>2</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>S 9</td>
<td>49</td>
<td>1</td>
<td>47</td>
<td>3</td>
</tr>
<tr>
<td>S 10</td>
<td>48</td>
<td>2</td>
<td>50</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 2: True Voice call attempts success count

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Successful Entry (True Acceptance)</th>
<th>Unsuccessful Entry (False Rejection)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S 1</td>
<td>45</td>
<td>5</td>
</tr>
<tr>
<td>S 2</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>S 3</td>
<td>45</td>
<td>5</td>
</tr>
<tr>
<td>S 4</td>
<td>49</td>
<td>1</td>
</tr>
<tr>
<td>S 5</td>
<td>45</td>
<td>5</td>
</tr>
<tr>
<td>S 6</td>
<td>47</td>
<td>3</td>
</tr>
<tr>
<td>S 7</td>
<td>44</td>
<td>6</td>
</tr>
<tr>
<td>S 8</td>
<td>48</td>
<td>2</td>
</tr>
<tr>
<td>S 9</td>
<td>46</td>
<td>4</td>
</tr>
<tr>
<td>S 10</td>
<td>48</td>
<td>2</td>
</tr>
</tbody>
</table>

The total Storage/memory and processing time is summarized in Table 3.

Table 3: Storage and processing time for different classifiers

<table>
<thead>
<tr>
<th>Classifiers</th>
<th>Storage Location</th>
<th>Training Time</th>
<th>Speaker Identification Time</th>
<th>Speaker Verification Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>VQ</td>
<td>1.0Mb</td>
<td>8.40s</td>
<td>15.75s</td>
<td>0.16s</td>
</tr>
<tr>
<td>DTW</td>
<td>4.0Mb</td>
<td>0.00s</td>
<td>0.80s</td>
<td>0.02s</td>
</tr>
<tr>
<td>HM</td>
<td>5.2Mb</td>
<td>250.0s</td>
<td>1.23s</td>
<td>0.02s</td>
</tr>
<tr>
<td>ANN</td>
<td>0.3Mb</td>
<td>1400.s</td>
<td>13.40s</td>
<td>0.26s</td>
</tr>
</tbody>
</table>

The training time listed is for each enrolment session. The speaker identification time was calculated on assumption that there were 100 enrolled users. From Table 3, the storage requirement needed for the VQ implementation was the least, with the DTW implementation required larger storage area.
The VQ implementation requires a comparatively moderate amount of memory. The VQ consumes less memory than the DTW, which was expected due to the lousy compression nature of the VQ implementation. All the classifiers evaluated required memory location which was easily made possible in current design.

The time needed to enrol a user varies drastically between the classifiers. The DTW implementation required 0.50 second for training and found to be acceptable, and may be used for online training. A person can be made to wait during an enrolment session, and thereafter the trained database may be verified. If the verification is unsuccessful, speech samples may be prompted again from the user to retrain the user database. The training time of VQ was well beyond the waiting time for a user who was enrolling. The training may be carried out offline, during the idle processing time of the voice recognition system.

The speaker identification time for DTW classifier was within acceptable limit. The identification time of the VQ was quite long and may not be suitable in certain applications like telephone banking and telephone credit cards. The training time can be reduced by using a more powerful DSP. The time needed for all 10 speakers who enrolled in the speaker recognition system were recorded. Prior to training, all speakers were briefed of the training procedure. Average training time was noted at 50.0 seconds. This included the voice sampling time of a minimum of 16.72 seconds. Sampling time increased due to verification of digit and login name sample. Speakers were requested by the system three times, if verification failed.

The average speaker recognition time was noted at 7.80 seconds. This timing included the prompt and sample time of 5 seconds. The system was able to make access decision in an average of 2.80 seconds after the voice sampling was completed.

5. CONCLUSION

There were many cases of security, financial transactions and personal data privacy breaches and fraud on the increase, the need for highly secure identification and personal verification technologies is becoming apparent. Therefore, in order to aid forensics in criminal identification, authentication in civilian applications and for preventing un-authorized access, there is a need to develop a voice recognition system that would be able to provide solutions for confidential financial transactions and personal data privacy that would reduce the high-tech computer theft or fraud in terms of access control, telephone banking and telephone credit cards.

This paper presents a model for maintaining data security and authenticity in voice-driven system whereby a system designed consists of memories and data acquisition modules that were well suited for a voice recognition system. Voice as a special characteristic of an individual, a form of biometric feature, could be used as a form of personal system identification and verification, and is recommended to be part of feature to be captured in the on-going government’s activities like the acquisition of National Identification Number, Drivers Licence, International Passport, Integrated Payroll and Personnel Information System (IPPIS), etc.

REFERENCES


E-Learning Functional Model: A Technology-Based Teaching Method For Providing Access To Sustainable Quality Education.

V.I. Osubor & S.C. Chiemeke
Department of Computer Science
Faculty of Physical Sciences
University of Benin
Benin City, Nigeria

ABSTRACT

Technology is driving radical change in the learning communities as traditional forms of teaching and learning are converted into internet or Web-Based environments. With the advent of the new communication revolution, the world is witnessing an expansion in learning. This new information revolution has enabled academic institutions to provide a flexible and more open learning environment for students. However, many countries and high school in Africa have not taken advantage of this revolution thus, traditional learning system is still the order of the day in many high schools today. In this study, we consider a technology-based teaching method called electronic learning (e-learning) that serves to disseminate a standard up-to-update version of a wide range of resources to a large number of students. A functional e-learning model is provided to x-ray how access to sustainable quality education can be achieved with technology based method.

Keywords: E-learning, Traditional Learning, Electronic Learning.

I. INTRODUCTION

With the advent of the new communication of revolution, the world is witnessing an expansion in learning. This new information revolution has enabled academic institutions to provide a flexible and more open learning environment for students. Harasim [7] points out that the convergence of new information technologies such as telecommunications, computers, satellites, and fibre optic technologies is making it easier for institutions to implement distance education. In Nigeria, distance education system is being considered as the most viable solution to the menace of satellite campuses recently banned by the Federal Government. The Government of Nigeria have been reactivating the National Open University of Nigeria (NOUN) that will be innovative and involving state-of-the-art teaching and learning multimedia packages. In [5], it was reported that the ultimate trend to the reactivation of the National Open University of Nigeria (NOUN) is towards delivery of courses through electronic networks.

African Universities and high schools have been confronted with numerous changes in their external and internal environments. They are forced to respond to emerging challenges such as the continual development in ICT and in recent years, the rapid decline in educational standard is evident due to crumbling infrastructure, unpredictable academic calendar, flight of researchers and professional abroad, and the declining respect for its graduates across the globe.

Providing access to quality education for every student in the universities is a significant task that must be accomplished so that they can reach their potentials. Therefore, they must follow the prevailing trend in most parts of the world by applying new technologies to overhaul and enhance its educational materials and resources. Though these universities and high schools are confronted with critical challenges to meet the new demands for the ever increasing student growth, they lack expansion in terms of educational resources to accommodate the increasing number of students. Therefore, they need educational environments that would make them more responsive to the confronted challenges.

One of such environments and ways of providing access to educational needs is the electronic learning which helps to provide students with the opportunities to have access to available expert, best resources and up-to-update information. E-learning is fast becoming an accepted and indispensable part of the mainstream of educational systems especially in developed countries. This growth has been structured in the path by the internet among educators and trainers in the use of new internet–based and multimedia technologies. It is fast transforming the educational sector in the industrialized world and it is worthwhile considering how these new technology can be successfully implemented in many African high schools and universities thus this study x-rays the importance of e-learning system and educate high school on how it can be successfully implemented to provide access to sustainable quality education.
1.1 The E-Learning System

E-learning could be defined as learning processes that occur using electronic medium. E-learning as an education via the internet, network or standalone computer. E-learning is delivered using electronic gadgets such as computers, tablets, mobile phones, internet technology, etc. It is a unifying term that describes the fields of online learning, web-based training, and technology-delivered instructions. Other common names for e-learning include: Computer-Based Learning (CBL), Web-Based Learning (WBL) and Multimedia-Based Learning [4].

Taylor [9] reports that e-learning is not intended to be blended with other training delivery methods to increase overall learning effectiveness and reduce cost. Aber (2003), reported that e-learning provides a fast and effective way to use digital technologies to deliver needed knowledge. It takes two forms, both of which are independent of location, so you can access them directly from your office. One form is “self-paced learning”, in which training is conducted at your own discretion and convenience. The other form is “real-time audio/visual communications,” which allows you to access demonstration, seminars and so on as they occur.

Figure 1: Pictorial definition of an e-learning system
Source Authors’ Fieldwork, 2006

[6] clearly identifies the differences between e-learning and online learning. E-Learning represents the whole category of technology-based learning, while online learning is synonymous with web-based learning. In this case, online learning is actually a subset of e-learning. E-Learning (also called Technology-Based Learning) covers a wide set of applications and processes, including computer-based learning, web-based learning, virtual classrooms, and digital collaboration. In this study, we define e-learning as delivery of content via all electronic media including the internet, intranets, extranets, satellite broadcast, audio/video tape, interactive TV, and CD-ROM.

Terms like e-learning, technology-based learning, and web-based learning are defined and used differently by different organizations and user group. Moreover, use of these terms is constantly changing, as the world of e-learning evolves. Online learning (also called Web-Based Learning) constitutes just one part of technology-based learning (e-learning) and describe learning via the internet, Intranet and extranet. The terms online learning is used synonymously with web-based learning or internet-based learning in this report.
Comparisons Between E-Learning (EL) and Face-To-Face (F2F) Teaching (Traditional Learning)

Table 1: Comparison of E-learning and Traditional (Face-to-Face) learning in terms of Learning material, Learning process, Learner’s motivation and performance. Also, the process of discussion, the learning location and instructor’s role are considered in the comparison.

<table>
<thead>
<tr>
<th></th>
<th>E-Learning</th>
<th>F2F/Traditional Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning material</td>
<td>The learner is more actively engaged with the learning materials.</td>
<td>The learner is less actively engaged with the learning material.</td>
</tr>
<tr>
<td>Learning process</td>
<td>Most of the learning processes occur in groups or by the individual learner.</td>
<td>Learning is conducted with the whole class participating; There is almost no group or individual study.</td>
</tr>
<tr>
<td>Classroom discussions</td>
<td>The learner talks at least as much as or more than the Instructor.</td>
<td>The Instructor usually talks more than the student.</td>
</tr>
<tr>
<td>Collaboration and peer discussions</td>
<td>Allows for greater problem discussion, peer collaboration and synchronous discussion with the instruction; Allows for easy online collaboration, sharing &amp; networking of ideas between and among the learners.</td>
<td>Uneasy online collaboration, sharing and networking of ideas between and among learners, and discussions with the Instructor.</td>
</tr>
<tr>
<td>Subject matter</td>
<td>The learner is involved in determining the subject matter since studying is based on various information sources, including web data banks and net experts location by the learner; Subject matter is richer and includes materials in different formats.</td>
<td>The lesson is conducted by the Instructor based on the study program and existing curriculum; The subject matter is far from the learners.</td>
</tr>
<tr>
<td>Emphasis on research study</td>
<td>The learner learns “how” and less “what”, with the learning involving research study which combines searching for and collecting information from web databanks and authorities on the communications network.</td>
<td>The learner learns “what” and not “how” with the learner’s and the instructors busy completing the required subject matter quota; Learners are not involved in inquiry-based education and problem-solving.</td>
</tr>
<tr>
<td>Learner’s motivation and performance</td>
<td>The learners’ motivation is high as a result of their involvement in matters closer to them and the use of technology; Higher performance and sources for learners.</td>
<td>The learners’ motivation is low, because subject matter is “distant” from them; Performance of the learners are not as high as in e-learning.</td>
</tr>
<tr>
<td>Learning location and instructor’s role</td>
<td>Learning takes place with no fixed location (anywhere) and the instructor directs the learner to the information.</td>
<td>The learning takes place within the classroom and the school’s compound, with the instructor being the sole authority.</td>
</tr>
<tr>
<td>Lesson stricture</td>
<td>Structure of the lesson is affected by the group dynamics</td>
<td>Structure of the lesson and division of time is dictated by the instructor.</td>
</tr>
</tbody>
</table>

(Source: Authors’ Fieldwork, 2006)

2. E-LEARNING SYSTEM FUNCTIONAL MODEL

Figure 2 gives a visual representation of the components that make up an e-learning environment and the objects that must be moved among these components. It is a simple functional and conceptual model designed in this study to understand how the different components of an e-learning system might work together. [10] defines learning objects as digital resources that can be reused to support learning or instructional components that can be reused a number of times in different learning contexts. A learning offering is defined as content that is assembled into a package of learning (ideally including assessment components) that is then offered to learners as a unit.

An offering catalog may associate offerings with learning paths that lead to degrees, certifications and/or skills. This catalog may be integrated with a more general content repository or may be a separate component, depending on the physical architecture of the learning environment. Content repositories are storehouses for learning objects and can be accessed on one hand by people and systems creating content, and on the other hand by people and systems using the content. Suffice it to say that content repositories must maintain a searchable index of learning objects and ideally, descriptive information about the structure and the properties of the objects.
Content (and Assessment)  Authoring tools and services allow subject matter experts and instructional developers to create and modify learning content objects. Professional Instructional Developers typically require these tools to provide a rich set of functions, whereas subject matter experts are better served by tools that are easy to use and learn, and provide standard templates for the content being developed. It is important for authoring tools to allow content authors to locate existing content to reuse or repurpose it rather than completely recreate it thus, in an ideal e-learning environment as in this study, authoring tools integrate smoothly with content repositories, allowing them to find, retrieve, modify, store, and replace objects and their metadata. Content Assembly refers to the linking of content objects together into cohesive learning modules with navigation between objects clearly defined and with assessments associated to appropriate content. Rules and/or behaviours for navigation through an offering is established during content assembly. It is usually performed using a different tool than the authoring tool used to create the learning objects, although many authoring tools include assembly capabilities. Thus, content assembly tools may support the creation and application of content templates that act as the basis for packaging content consistently and efficiently into learning modules.

Content/Catalog management is the process of defining the learning that will be offered to different audiences, establishing learning plans (degree paths, certification paths, skill development curricula), scheduling the resources needed to support learning delivery, establishing the business processes for registering learners in offerings, and making the offering catalog accessible to the target audiences. Catalog manager components are typically, interfaces that allow authorized individuals to make learning available and to set access rules, restrictions, prices, and so on [3]

Learning is ultimately about learners, and thus, e-learning systems typically keep information about the learners that use them. This information includes: personal data, learning plans (e.g. degree plan), learning history, certifications and degrees, assessments of knowledge (skills and competencies), and the status of participation in active learning (registration, process). The sum total of this information is called a “learner profile”, and e-learning systems require a component that manages this profile. The Learner Profile Manager makes learner info available to other components and retrieves and updates learner information on the basis of data reported by other components. Each learner’s profile is stored in the “Learner profile Repository”. The Delivery Environment provides the learner with access to learning content and other components of a learning environment such as chat, email, quizzes, multimedia players, collaboration tools, application sharing, shared whiteboards, equation editors etc. The environment also provides tools for instructors if there is an instructor-led component of the learning. Data on a learner’s activities and status in an offering may be passed back to the learner’s profile.

2.1 Benefits of E-Learning System in Education

The benefits of e-learning to both the Instructor/Teacher and the Learner/Student are numerous. Some of the advantages of e-learning to the Instructor include: Reduced overall cost since cost associated with instructor’s salaries, meeting room rentals student travel, lodging, meals etc are eliminated; Kruse [8] points out that learning times are reduced by an average of 40-60% through e-learning; expert knowledge is captured and communicated; essential element of training and initiative can be automated, and consistent content delivery is possible with asynchronous, self-paced e-learning. [1] and [2] noted based on his findings that e-learning has created an operational improvement in the learning process and also improved the quantitative and qualitative changes in the learning process.

The benefit of e-learning to the learner include: increased retention; reduced learning time; students can complete training conveniently at off-hours or from home based on on-demand availability; reduces stress and increases satisfaction for both slow and quick learners (since it allows self-pacing); interactively engages users, pushing them rather than pulling them through training etc. Also Kruse [8] reports that e-learning reduces burden of responsibility of mastery or confidence than refresher or quick reference materials makes available.

3. DESIGNING A COURSE FOR E-LEARNING SYSTEM

There are two major e-learning delivery modes used in delivering materials to students. They are:

i. Synchronous Delivery Mode: This mode of delivery allows real time interaction between both parties (students and instructor). This form of delivery could be in form of multicast web seminars, chat or even tele-video conferencing. This is an efficient e-learning delivery channel because student could ask question instantly and the instructor would be able to carry everybody along

ii. Asynchronous Delivery Mode: This mode of delivery does not allow real time interaction between the student and instructor. This form of delivery channels could be through e-mail, CD-ROMS. Instructors and Course Developers must simultaneously consider the following when or before designing an e-learning course to ensure optimum impact:

iii. Course Material: Instructors need course contents that are current and relevant to the student’s career choices thus, they need to know the students’ needs. All instructors must work to keep up with current course content since these contents are dynamic.
iv. **Student’s learning styles:** Instructors should plan course presentations based on the students learning styles and not based on his/her learning style or the way he/she was taught.

v. **Methodologies of presentation:** Methods of effective presentation can widely vary among disciplines thus, the right choice of technology suitable for enhancing different learning methods should be chosen.

vi. **Assessment:** This is considered at the development stage of the course material. At the beginning of the course, it is very important to know the beginning level of students understanding of the subject. It is also important that all subjects know their own learning styles and the best way to approach learning based upon them.

However, the basic building blocks necessary for the successful implementation of a sustainable e-learning system include:

- Highly available and quality broadband internet connectivity.
- PCs and Workstations with at least 450MHZ speed with speakers, soundcard, printer, Operating System (Windows95, Windows98, Windows2000 or WindowsXP) and RAM size of 128MB or more.
- Reliable e-mail software and active e-mail account.
- MODEM speed of 56Kbps (Kilobits per second)
- Reliable electric power supply available 24-hours a day and 7days a week.
- UPS and backup power systems such as those that include solar power batteries and wind energy.

### 3.1 Barriers to implementing e-learning in African Schools

i. **Technological and Cost Infrastructure:** E-Learning requires a reasonable technological infrastructure and cost and e-learning can happen whenever this is lacking. They include: setting up servers, PCs and Internet at reasonable access speed of at least 512kps. Also, issues of reliance, maintenance, quality of service, IT support such as learning technologist or e-moderator are required.

ii. **Social aspect:** Learners tend to feel isolated. Trainers are worried that they will lose their jobs. Learners and Trainers need to pick up skills for online learning and training.

iii. **Pedagogical aspect:** It is imperative to familiarize learners and trainers with new way of learning because education will become learner-directed than instructor-directed. Learners need to discipline themselves, and learn a new way of learning, gathering information, getting resources and sharing knowledge and experience with others.

iv. **Mindset aspect:** E-Learning requires a total change of mindset thus, strong support from senior management is extremely important for any institution that wants to implement e-learning. However, this is lacking in many Africa High Schools.

### 3.2 Why Every High Schools in Africa Should Embrace And Implement E-Learning

From our study, we have seen that the potentials of e-learning system cannot be fully realized until staff and students are using it. Therefore, every high school in Africa should embrace and implement e-learning for the following reasons:

i. It is accessible 24hours a day, 7days a week and helps improve learner’s effectiveness and retention levels.

ii. It can serve as standalone course for refresher training, pre-class preparation, corrective training, as an assessment tool and for just in-time performance support.

iii. It provides greater and more equal access to quality education. Training can be delivered to places where student populations are not significant enough to support an instructor or where qualified teachers are unavailable.

iv. Through e-learning, students are given greater control and responsibility for their own learning processes as they discover new areas of interest and accomplishment. They become knowledge-seekers and not just instructor-recipients, and actively construct it on the basis of their prior knowledge and experiences, which the traditional learning most of the time simply ignore.

v. E-Learning seeks to improve the quality of graduates by utilizing modern instructional materials and methods, including increase in the use of ICT in teaching and research. Also, considering the increasing number of intakes into our Universities and high schools which in most cases, is not matched with classrooms space and structures, e-learning will go a long way to improve access to quality education by students.
4. CONCLUSION AND RECOMMENDATION

Sustainability of e-learning is dependent on the recurrent costs on information resources. The followings are recommended for every high school in Africa for sustainable quality education.

1. Every high school should provide adequate funds for the purchase of e-learning materials for use in all e-learning centres to continually upgrade staff skills.

2. In each school/college, an e-learning team consisting of a content expert, an instructional designer, a programmer, a web-interface designer, a graphic designer and an assessment expert adequately equipped with ICT skills, teaching skills (experience), operational and good communication skills, should be constituted.

3. All academic staff should be made to undergo compulsory training in education techniques with emphasis on e-learning. Also, all students need to be trained on a continuous basis to equip them with requisite e-learning skills.

4. Each department should set up an e-learning laboratory to develop local capacity in developing and evaluating appropriate training software.

5. All students are required to take the prescribed introductory level module(s) as a requirement for e-learning.

6. Appropriate common e-learning infrastructure and software responsive to academic needs should be established in all designated e-learning centers.

7. A helpdesk should exist as a point of contact, staffed by personal who can give immediate support or referral to specialists, with the e-learning team continually liaising with the helpdesk for information resources management.

8. Strategies have to be developed for collaborative partnerships, broadband connectivity, universal access, e-learning and e-administration.

9. The university’s policies should create organizational (trainer capacity and training management) and technical (practice laboratory and computer-based training tools, self-paced training mode) conditions assuring continuous in-house e-learning capabilities in the long term.

REFERENCES


GOEmbed: A Smart SMS-SQL Database Management System for Low-Cost Microcontrollers

N.E. Osegi
Department of Information and Communication Technology
National Open University of Nigeria
Lagos State, Nigeria
E-mail: nd@osegi.com
Website: www.osegi.com
Tel: +234 7030081615

P. Enyindah
Department of Computer Science
University of Port Harcourt
Rivers State, Nigeria
E-mail: probi2k3@yahoo.com
Tel: +234 8036710489

ABSTRACT

The era of the Internet of things (IoT’s), machine-to-machine (M2M) and human-to-machine (h2m) computing has heralded the development of a modern-day smart industry in which humanoids can co-operate, co-exist and interact seamlessly. Currently, there are many projects in this area of smart communication and thus giving rise to an industry electrified by smart things. In this paper we present a novel smart database management system (dbms), GOptimaEmbed, for intelligent querying of databases in device constrained embedded systems. The system uses genetic algorithms as main search engine and simplifies the query process using stored in-memory model based on an invented device dependent Short-messaging-Structured Query Language (SMS-SQL) schema translator. In addition, querying is done over the air using integrated GSM module in the smart space. The system has been applied to querying a plant database and results were quite satisfactory.

Keywords – GOptimaEmbed, smart dbms, genetic algorithms, SMS-SQL

I. INTRODUCTION

A smart dbms is one in which information storage and retrieval is done entirely by smart devices. Smart devices on the other hand are ubiquitous knowledge based devices with inferential abilities. Smart processing involves the intelligent operations, timely intervention, extraction and utilisation of abstract and real data in smart space or environment. The capability of a smart system is thus dependent on the quality and specifications of the smart devices employed which is largely influenced by the nature of the application and the environment. Typical smart devices used include from high-end and high cost microprocessors such as ARM Cortex M3, AVR 32-bit, and System-on-chips (SoC’s) to low-end and cheap microprocessors typically of the 8-bit category. Smart databases are useful when traditional databases may not be a reliable and flexible option as in real-time sensor-actuated information networks and dynamic databases.

When the database is not very large and memory requirements are small, a smart dbms may be preferable. Currently, improvement in smart device manufacture and design miniaturization has led to the available of very cheap high performance low-end or resource constrained microprocessor systems such as the Arduino [1]. Smart dbms systems can encourage and facilitate the end-user programming modelling. Smart systems generally refer to a class of expert system devices for which the level of abstraction is a variant of the input to the system. In smart solutions, two core requirements are essential:

- The knowledge or information and
- The Inferential Mechanism

In order to effectively manipulate such a system, there is need for the presence of at least two main players. We may refer to the actors or players in a smart solution as “humanoids”. Here humanoids” is an acronym that stands for machine-man cooperation or human-robot interface systems.
Currently, smart systems employ high-end microprocessors and embedded systems such as ARM, FPGA, AMD and advanced computation libraries. These architectures may employ a subset of SQL called tinySQL [2] or SQL lite as well as a client-server engine for database management [3]. However, little work has been carried out by researchers on the potential of SMS-SQL database integration in low-end devices. The lack of tight integration in existing systems has also reduced the effectiveness of information retrieval from these systems. This is due to lack of suitable and readily available SMS-SQL syntax translator for in-memory database and an adaptable and robust search algorithm for devices of the low-end category. Since this devices have a useful role to play in the smart world, dbms strategies have to be devised to tackle this category of embedded processors. GOptimaEmbed, is an alternative smart dbms using the strength of a genetic algorithmic search model and an efficient SMS-SQL syntax translator for in-memory database queries. The core components of this system will be described briefly here to enable the reader gain an insight into the “world of smart intelligence”.

1.1 Genetic Algorithms
Genetic algorithms (GA) introduced by Holland [4], are bio-inspired artificial (or machine) learning search procedures developed to find exact or near solutions to a variety of optimality and knowledge discovery problems [5]. The GA’s are actually computer programs that reside on a microcomputer device and offer the benefits of bio-inspired artificial reproduction which include random mutations, cross-over, selection and recombination from a parent population, as well as fitness measures and checks. The artificial reproduction is mathematically modelled using natural heuristics algorithms and decision trees with a stopping function when fitness is met. Fitness is actually described as a core mathematical function or logic operation that must be met for an individual to be deemed successful. Artificial reproduction gives rise to offspring’s who later become parents themselves after a successful survival (fitting) exercise. These reproduction and fitting process is carried out over many generations. In practice the generations are typically set to a finite number say 10 to 100 for embedded systems and small to medium-scale AI based software projects. The fitted individuals are successful and are said to be potential solutions to the problem. The population in a GA process is made up of individuals also called chromosomes (genes) who participate in the evolutionary process. All the genes in a given population give rise to the gene pool. The Genotype describes the individuals’ structure and quality. For a genotype, a defining length is given which is typically of fixed length strings or integer values representation and from which a phenotype may be decoded – the decoding process is practically done using special array structure handling. The phenotype contains the alleles (or gene values). They serve as input values that must be fed to the fitness function to validate the evolution process. The gene locus or location of a gene on a chromosome will be dependant to a large extent on the nature of the cross-over scheme and to a small extent on the type of mutation. Several mutation and crossover schemes have been given in [6] and have been successfully applied in practice. In a time dependent manner a selection is made using any of the selection schemes of roulette wheel, tournament or truncation described in [6, 7, 8]. The advantages of a GA approach is that they can serve as very good approximate reasoners and converge to the local minima quickly and effectively. Thus, through the concept of natural evolution- mutations, crossover, recombination, a GA optimized solution can possibly attain the solution state earlier than conventional systems.

Some GA’s for low end microprocessors have been developed in [5, and 9]. However, these algorithms are needlessly complex, and not flexible for database integration. In this regard we seek to implement an effective but simple algorithm on a small footprint microcontroller using a modified GA optimisation scheme adapted from [6]. This has the advantage of a minimum barebones approach to the Genetic search problem and can easily be integrated into a C++ library class for further flexibility.

1.2 Statement of Problem
In recent times there has been a rise in dedicated ubiquitous smart microcomputer products for information retrieval and storage. However, most of these products are either too expensive or are not sensor and human friendly. They lack interactive and controllability features. By using high-end databases, they are generally non-customizable making them resemble their PC counterparts. Also, the potential of remote query exploitation using the SMS-SQL approach has not been fully explored in the area of device constrained smart computers.  

1.3 Research Objectives
Our research objectives are two-fold.

First, we will build a smart computer model that will implement a real-time smart dbms for device constrained devices.

Secondly, we will validate the model using the SMS-SQL queries as a proof-of-concept

1.4 Embedded Processors/Systems
Embedded processors are typically microprocessors with a single CPU that emphasises controllability rather than complex computational arithmetic. However, modern day embedded processors can carry out some level of complex arithmetic with some reduction in performance. In some cases these processors are referred to as microcontrollers or single-board-computers. Examples of such include the 8051’s, AVR series, ARM Cortex, Arduino, Intel Galileo, and variants thereof. Embedded processors form the heart and soul of an embedded system and provide the foundation for intelligence building in a smart product. In our prototype, we have used the Arduino microcomputer which is based on the ATMega328P, an AVR 8-bit core microprocessor as a representative processor.
1.5 GSM Modules

GSM modules are communication devices that come with a full instruction set for interfacing embedded processors. These devices come in various sizes and shapes and allow for remote messaging, internet and voice services via a GSM network. Some typical examples include the Quectel M10 GSM Module, SM5100B Module and the GSM click. We have utilized the Quectel M10 GSM Module since this feature an onboard antennae and integrates the Arduino shield concept.

1.6 Device SQL and In-Memory Model Databases

Device SQL refers to SQL for an embedded processor database. This database is often of the Array structure and builds on the internal memory of the device in question. The benefits of Device SQL cannot be overemphasized and has been described in [10]. In designing a Device SQL framework one needs not bother about implementation language, technology or the issue of database drivers. The only main drawback of such systems is that most low-end microprocessors have limited memory constraints and speed of processing. However, this can be worked around by modifying the data structures with strict types. Since the focus is on micro-data and not Big Data, this shouldn’t be a problem. Currently modern low cost microprocessors that offer high memory in megabytes range which can support both processing high volumes of image and text are available. With the DeviceSQL concept, the embedded processor should be responsible for its own dbms requirements and only call upon external memory when its internal memory is not sufficient.

In this paper we emphasise on device dependent database since our storage requirements is small and when the need to adapt the system to respond to dynamic inputs. We call this small data “micro-data”. This typically takes a few kilobytes and is desirable particularly in Sensor actuated networks.

2. RELATED WORKS

Very little work has been done in the area of smart database query SMS-SQL integrating Artificial Intelligence schemes like GA’s. In [11] an SMS-SQL has been proposed for the management and querying of a database. The features of the existing system include a high-end network processor and an Operating System (OS) based on the Unix kernel in a client-server environment. The system require additional component such as USB Modem and two ports and this can increase the overall cost of the system. Using a PC-based client-server dependent protocol – see Figure 1, the system still requires traditional database connectivity for its operation. In addition the choice of non-native language (they used Perl language) means increases in cost of memory allocated. Their system thereby suffers from the database connectivity (dbc) bottleneck.

MGuide:

This is a System for mobile information access and is based on Java 2 Microedition (j2me) standard [12]. It supports both static and dynamic information access. The system requires two j2me compliant cell-phones for operation. One of the phones acts as a server connected to the internet through PC. The other is the user requesting for information. As stated earlier, MGuide use j2me which has the drawback of lack of support for retrieval of messages from inbox directly from the screen. Another drawback of the system is the use of traditional database which cannot be implemented on low end microcontrollers.

TinyDB:

TinyDB [13] is a fully functional SQL equivalent small footprint open-source database for sensor actuated networks using battery-powered nodes or motes (sensor microprocessors with integrated wireless radio). It has been successfully deployed in low-end micro-computing platforms. It typically runs on motes from the Crossbow Technologies and compiles using the nesC compiler [14]. However TinyDB does not accommodate external queries using SMS since it does not encourage the humanoid model. Nevertheless, it possesses key features that can make low-end microprocessors a success story. Its query syntax can easily be transformed and integrated in both embedded personal computing platforms and standalone systems.

ExtremeDB:

ExtremeDB [15] is a commercial dbms that also support small footprint database for embedded processors with in-memory database features as well as large PC database. It runs on embedded real time operating systems (RTOS), as well as traditional operating systems like windows and linux. No support for SMS-SQL translation has been explicitly defined in their model.

To implement a low-end microprocessor based SMS-SQL dbms we need to employ simple data structures with emphasis on simplified device SQL query definitions, SMS information processing and improved search algorithm such as genetic algorithm. This can be made possible using the GOptimaEmbed framework.
3. PROPOSED SYSTEMS MODELLING AND DESIGN

One major challenging in devising an SMS-SQL translator is decoding the required message in microprocessor dynamically, parsing message to data handlers and making the report in a timely and efficient manner. In the design of our systems model we envisage two use cases

1. Static In-memory database (SID) Use-Case
2. Dynamic (Sensor) Actuated In-memory database (DID)

The *GOptimaEmbed* model follows from the need to realize real-time data acquisition solutions that will facilitate information retrieval from small data repositories (or libraries) as well as real time data from production stations. Since it is hardware as well as software system, it conforms the class of computing referred to as “Physical Computing” [16]. Our software project/modelling methodology shall encourage the functional-object-oriented paradigm with an emphasis on rapid prototyping. We shall focus on this aspect of software engineering based on Physical Computers.

a. Systems Protocol

The protocol of the proposed system is shown in Figure 2. The SMS is transmitted over GSM network to remote server. The server queries the specified database for requested information and reports back to the user the extracted information.
b. Systems Architecture
The architecture of GOptimaEmbed describes a modular component system with parallel processing channels for simultaneous query retrieval. Concurrency is achieved using multi-tasked time schedule function blocks. To minimize energy consumption the system activates power down mode during inactivity. The system is useful as queries can be redirected to as many repositories as possible using microprocessor firmware. The system is characterized by an applications development framework consisting of GSM class libraries, GSM driver, Network Interfaces, the hardware components, a GA Optimizer and the DeviceSQL In-memory Translator functions. An Arduino Bootloader is included as this provides the enabling platform and mechanism for the system to be implemented, operate and function without necessary any specialized source code downloader hardware [1]. The systems component diagram is shown in Figure 3 below.

![GOptimaEmbed Component Architecture](image)

**Fig 3: GOptimaEmbed Component Architecture**

c. SMS-SQL Query Process Model
The SMS-SQL Query Process Model describes the message format for requesting information from a resource. In Figure 4 is depicted the sequences of states that must be overcome for a smart dbms access in a finite time. This process can generally occur in two stages:

i. Querying Structure

ii. Query Translation and Parser

![Query Process Model](image)

**Fig 4: Query Process Model**
d. **SMS-**
   i. **SQL Query structure**

   Users can send query via sms using simple syntax given as:

   Query:  db1 tb1 at1 va4

   Where,
   - db1 stands for database 1
   - tb1 stands from table 1
   - at1 stands for first attribute and,
   - va4 stands for attribute value of 4.

   The SQL equivalent is given as:

   ```
   Select Tb1 from Db1 where At1 is = “4”
   ```

   This SQL restructuring is necessary to facilitate the end user query entries and to simplify the translation at the Server end.

   ii. **SQL- translation**

   The translation process involves decoding the necessary parameters from the received query string and discarding irrelevant details. This is done by employing suitable string indexing functions and equivalency commands. The Boolean logic of if-else decision is then used to extract message tokens from the requisite data stores in-memory. This translation process is supported by a genetically optimized search routine the modifications of which is given in Appendix I.

   e. **Design**

   The design will typically include the specifications, a block level description of the system and the algorithmics. The system is based on the Arduino single-board microcomputer, a powerful and cost-effective prototyping platform for resource constrained microprocessor systems. Smart communication by the humanoids is effected by using the shield concept where a smart communication module is pin-to-pin structurally compatible to the Arduino. The block design is shown in Figure 5.

![Fig 5: Systems Block Design of GOptimalEmbed](image)

---

**i. System Components**

The initial prototype includes the power supply unit, GSM shield unit (GU), and embedded microprocessor system unit (MSU), and indications unit (IU). The embedded MSU uses the Arduino framework based on the Atmega328P RISC microcontroller/processor. IU gives status of queried species using light emitting diodes (LED’s). This indicationii. is used for real-time debugging purposes and may be excluded from the final prototype.

**ii. Specifications for Smart dbms**

The specifications include the minimum hardware specific requirements, the software and the database requirements. The algorithmic requirements shall include relevant points as described in [17]. Detailed specifications for smart microcomputer and communication module hardware can be obtained from [18] and [19] respectively. Sample data from the Iris-dataset has also been obtained from [20].

**Hardware Requirements:**

- Power supply – 2A; 5v regulated power pack or 5v:6AH battery
- Embedded Processor – Arduino
- Current draw: 2.0A max burst
- Voltage Tolerance: 7.0v-12.0v +/- 3%
- GSM Module (Shield) – Quectel M10 Module
- Current draw: 2.0A
- Voltage Tolerance: 4.5v +/- 3%

   (If using GSM module same voltage range should apply as in item II)
- LED indicators – Visual Interactive Species Labels
- Red – Iris-Setosa
- Orange – Iris-Versicolor
- Green – Iris-Virginica

**Software Requirements:**

- IDE – Arduino 0023
- Number of Attributes – 3
- Attributes – Sepal-length, Sepal width, Petal length, Petal width
- Key Attribute – Sepal length

**Algorithmic Requirements:**

- Core Input Cardinality – 4 (Platform, Population size, Number of Generations and objectives)
ii. Topology – function and procedure calls (no classes)

iii. Time Complexity – Logarithmic, linear

iv. Refine criterion – Worst-case

v. iTest – Turing Test

iii. Simplifying Algorithm
The simplified algorithm using GOptimaEmbed, integrates the SMS-SQL translation with a Genetic Optimizer. It is possible to modify the fitness function of the optimizer to account for equivalency-to-zero checks; less-than or equal-to checks; and complex function checks either singly-or initii. a multi-objective manner. Checks can also be made against dynamic sensor actuated inputs. Combination of various checks in various orders is also possible.

The steps taken for smart database query and report are as follows:

Step1: Initialize working variables: database store, population size, number of generations, insertion and db storage variables, classes and library headers

Step2: Set Generations Counter

Step3: Initialize population

Step4: Define db_check //Function that extracts //relevant message report from requested database // and sends to Pre-defined Client ID’s.

Step5: Query (GSM)

Step6: Decode SMS strings

Step7: Translate SMS to Integer

Step8: Fitness Checks

Step9: Fit = \text{abs}(\text{selected individual} - \text{desired individual}) //desired individual is allele in SMS //Message Criterion: == 0; <=threshold //typical threshold = 1;

vi. If (fitted)

vii. Extract alleles from db storage

viii. String/Integer match

ix. Call (db_check)

x. Print Fit

Else : GoTo Step 5

xi. Return Fit

Step10: Select Fitted Individuals

Step11: Cross-over Reproduction

Step12: Uniform Mutation

Step13: Return to Step 5

End

4. IMPLEMENTATION AND RESULTS

Figure 6 shows a prototype outlook of the smart system with GSM module shield interfaced to the Arduino smart computer. With the shield concept the developer can concentrate more on the implementation logic rather than on product design and thereby reducing overall system development time by half.

Fig 6: GOptimaEmbed Smart Module Interfacing

The results of simulation for various queries on the smart dbms are given in Table1. The difference between \( t_{sent} \) and \( t_{received} \) gives the latency which is about 2s. Users of GOptimaEmbed can send query using sms in the form given in section 3.3.1. The graphic report may vary depending on phone make and model.
Table 1: Query Report

<table>
<thead>
<tr>
<th>Query No.</th>
<th>SMS Query</th>
<th>t_sent (s)</th>
<th>t_received (s)</th>
<th>Query Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>dbiris tbiris atsepl va2</td>
<td>19:11p.m Feb 8</td>
<td>19:12p.m Feb 8</td>
<td>Species Found Iris-Setosa</td>
</tr>
<tr>
<td>2</td>
<td>dbiris tbiris atsepl va2</td>
<td>19:49p.m Feb 8</td>
<td>19:50p.m Feb 8</td>
<td>Species Found Iris-Setosa</td>
</tr>
<tr>
<td>3</td>
<td>dbiris tbiris atsepl va8</td>
<td>19:19p.m Feb 8</td>
<td>19:21p.m Feb 8</td>
<td>Species Found Iris-Virginica</td>
</tr>
<tr>
<td>4</td>
<td>dburis tbiris atsepl va6</td>
<td>19:19p.m Feb 9</td>
<td>19:21p.m Feb 9</td>
<td>Species Found IVS</td>
</tr>
</tbody>
</table>

5. CONCLUSIONS, RECOMMENDATIONS AND FUTURE WORK

Smart dbms using GOptimaEmbed is a cost-effective way of implementing real-time smart solutions for information retrieval and storage in resource constrained devices. When optimized with GA’s, such systems can help traditional data storage and retrieval search systems make good decisions on the desired information when faced with in-exact, fuzzy or approximate requests. Smart dbms’s can also revolutionize the information engineering process and improve the fortunes of GSM operators, aid educationists and students, and anyone looking for cost-effective and ubiquitous ways of accessing their valuable information anywhere, anytime in an intelligent way. We recommend this tool to the Nigerian Communications Commission or similar body as a useful test instrument for assessing the relative performance and quality of GSM operators in the country. In future, we plan to integrate more SMS-SQL like functions into GOptimaEmbed and apply it to dynamic sensor networks for real time monitoring of plant species growth profiles.

REFERENCES

APPENDIX

Source-Code Listing for Implementing *GOptimaEmbed*

using Arduino Framework:

```c
# Class libraries
#*/
#include <String.h>
#include <ClientGSM.h>
#include <GSM.h>
#include <cinetGSM.h>
#include <LOG.h>
#include <QuectelM10.h>
#include <ServerGSM.h>
#include <Streaming.h>
#include <UDPGSM.h>
#include <WideTextFinder.h>
#include <EEPROM.h>
#include "QuectelM10.h"
#include <NewSoftSerial.h>

// Variables and Constants
const unsigned int pop_bounds[2] = {4, 8};
const unsigned int pop_size = 16;
const unsigned int num_gens = 40;
const unsigned int num_muts = 1;
const unsigned int def_length = 0;
unsigned int min_bound = 0;
unsigned int max_bound = 0;
unsigned int ipop[16]; // Initial Population
unsigned int ipop_max = 0;
unsigned int fit_pop[16];
unsigned int cfit_pop[16];
unsigned int j[16];
unsigned int fit;
unsigned int sms_in;
const unsigned int td = 5000;
unsigned int i;
unsigned int stop_criterion;
unsigned int total_fitness;
int sms_report = 0;
char smsbuffer[160];
char n[40];

unsigned int intx1 = 0;
unsigned int intx2 = 0;

void setup() {
    // randomSeed(analogRead(0));
    // Light Indicators
    pinMode(8, OUTPUT);
    pinMode(10, OUTPUT);
    pinMode(12, OUTPUT);

    // Open serial communications:
    Serial.begin(9600);
    Serial.println("GSM Shield testing.");
    // Start configuration.
    if (gsm.begin())
        Serial.println("nstatus=READY");
    else Serial.println("nstatus=IDLE");

    min_bound = pop_bounds[0];
    max_bound = pop_bounds[1];
    def_length = (max_bound - min_bound)/pop_size;

    // send an floatro:
    // sms_in = 2;
    }

// FUNCTIONS USED/CREATED
int queryGSM() {
    }

gsm.readSMS(smsbuffer, 160, n, 20);
    delay(500);

    }

int dbcheck_1(String int_x1, String int_x2) {
    if (int_x1.equals("52") && int_x2.equals("57"))
        }
    switch (dbase_species[i])
        {
        case 4:
            gsm.sendSMS("07030081615", "Species found is:");
gsm.sendSMS("07030081615", "Iris-Setosa");
```

```c
```
```c
if(int_x1.equals("52") && int_x2.equals("69"))
{
    switch (dbase_species[i])
    {
        case 4:
            gsm.sendSMS("07030081615", "IS");
        break;
        case 5:
            gsm.sendSMS("07030081615", "IS");
        break;
        case 6:
            gsm.sendSMS("07030081615", "IVS");
        break;
        case 7:
            gsm.sendSMS("07030081615", "IVG");
        break;
    }
}
```
Serial.println("Population Attainable = " + String(ipop[i]));

// fit_pop[i] = 2*def_length*ipop[i]/ipop_max;
fit_pop[i] = abs(ipopn[ipop[i]]-sms_in);
Serial.println("Fitted Population = " + String(fit_pop[i]));
Serial.println("Point Locii i = " + String(i));

//cfit_pop[i] = fit_pop[i];
// fit_pop[i] = abs(fit_pop[i]-sms_in);
Serial.println("Absolute fit = ");
Serial.println(fit_pop[i]);

// Stopping Criterion
//stop_criterion = Geni;
if(fit_pop[i] <= 1)
{
    // Serial.println("cfit");
    //Serial.println(cfit_pop[i]);
    Serial.println("Solution found at Locus: " + String(i));
    Serial.println("waiting....");
    // gsm.sendSMS("07030081615","Species found is:");
    //gsm.sendSMS("07030081615", "Iris-Setosa");
    dbcheck_1(int_x1, int_x2);
    Serial.println("\nSMS sent OK");
    //break;
    break;
}
}

// Selection begins here:
// use roulette selection (-> need pos. fitness!)

// REPRODUCTION
// parents 2i-1 and 2i make two new children
// 2i-1 and 2i crossover
// use arithmetic crossover

for (i= 0; i<=pop_size; i= i+2)
{
    ipop[i] = random(pop_size)+parent[i]; // + (1-random(pop_size))*parent[i+1];
    ipop[i+1] = random(pop_size)+parent[i+1];
    // ipop[i] = random(pop_size)*parent[i] + (1-random(pop_size))*parent[i+1];
    // ipop[i+1] = (1-random(pop_size))*parent[i] + random(pop_size)*parent[i+1];
}

// Cross-over script ends here

// mutation
// use uniform mutation
for (i= 0; i<=num_muts; i++)
{
    ipop[random(pop_size)] = pop_bounds[0] + random(pop_size); /*def_length*/
    delay(1000);
}

// Selection ends here
Towards the Development of a Time-Out Multiple C-R CAPTCHA Framework Using Integrated Mathematical Modeling

J.O. Okesola  
School of Computing  
University of South Africa  
South Africa  
tunji_okesola@yahoo.co.uk

Longe O. B.  
Department of Computer Science & Mathematics  
Adeleke University, Ede, Osun State, Nigeria.  
longeolumide@fulbrightmail.org

A.P. Obi  
42d Earlham grove  
London  
E7 9AW.  
prcs_obi@yahoo.com

ABSTRACT

The internet has suffered from large forms of insecurity ranging from scamming, hacking and theft of information. Lately the use of CAPTCHAs has become a common security tool for authentication and authorization. However CAPTCHAs has suffered from certain vulnerabilities in the context of the simplicity offered by the challenge-response scenario and its timing which leaves room for improvement. This paper proposes a Time-Out Multiple Challenge-Response (C-R) CAPTCHA Framework that Utilizes Mathematical Modelling as a basis for overcoming some of the challenges faced by current CAPTCHA Systems. Our approach ensures security during the authorization and authentication process.

Keywords: Time-Out System, Challenge-Response, Security, Authorization, Authentication and CAPTCHA

I. INTRODUCTION

CAPTCHA stands for “Completely Automated Public Turing Test to Tell Computers and Human Apart”. They are slightly distorted images, challenges or tests administered automatically over networks that can distinguish between people and machines (automated script) and thus protect web services from abuse by programs disguising as human users. [2][4][5]. Captcha is a security poser administered over the internet to monitor the authenticity of each usage of resources on the internet. They are challenges or tests meant to be easily solved by humans while remaining too hard to be economically solved by computers. For example, humans can read distorted text as displayed in figure 1 but current computer programs cannot. Captcha are usually a single challenge response system but there are studies that present the double challenge response system called 2RC –(dualistic captcha 2010).

CAPTCHAs are used by Yahoo, Hotmail, PayPal, Microsoft, TicketMaster, Register.com, Google and many other popular sites to prevent and protect services against automatic script attacks and automated registrations.

Fig 1: Distorted Image Example
CAPTCHAs work because no computer program can currently read distorted text as well as a human being. Cyber security involves protecting information by preventing, detecting and responding to attacks through various mechanisms. Measures currently in use include Intrusion Detection Systems (IDSs) and Intrusion Prevention Systems (IPSs). These protective measures are to combat the risks involved in the application of computers and internet technology infrastructures. Other measures taken against threats to resources on the internet include the use of cryptography (decryption and encryption), password, signatures, iris recognition, Captcha, etc. [1].

2. CAPTCHAs IN PERSPECTIVE

The term CAPTCHA was coined in 2000 by Luis von Ahn, Manuel Blum, Nicholas Hopper and John Langford of Carnegie Mellon University [4][7]. Previous studies showed CAPTCHA as an alphanumeric single challenge response system. However, malicious attackers have been able to use automated script to answer the challenge with the aid of OCR (Optical Character Recognition) while sophisticated attackers employ social engineering (relay attacks) to address the problem. To be effective, a CAPTCHA must be difficult enough to discourage script attacks by raising the computation and/or development cost of been broken to an unprofitable level. At the same time, it must be easy enough to solve so as not discourage humans from using the service but at the same time maintain some good levels of security. (recaptcha.com). The single challenge system is such that the text pops in slightly distorted and the user is expected to retype the popped up text or audio on the space provided, in the case where the user fails to type in the text correctly, it changes the text and represents another text to the user and it goes on for as long as the network can take, the problem with this is that it takes up too much time that could be used for something else and in the case where the user is a non-human, it keeps allowing access to the site or information. Over the years there has not been a complete work on the remedy to this problem though work is on-going on some other ways to correct this anomaly and making the captcha more secured.[10][9].

2.1 Applications of CAPTCHAS.

Captcha is applicable to several practical security scenarios some of which are discussed in this section

Preventing Comment Spam in Blogs. Most bloggers are familiar with programs that submit bogus comments, usually for the purpose of raising search engine ranks of some website (wikipedia.org). By using a Captcha, one can ensure that only humans can enter comments on a blog. [14].

Protecting Website Registration. Several companies like Yahoo offer free email services. Until few years ago, most of these services suffered from a specific type of bot attack that would sign up for thousands of email accounts every minute [15]. The solution to this problem was to use captchas to ensure that only humans obtain free accounts. In general, free services should be protected with a captcha in order to prevent abuse by automated scripts. [9]

Protecting Email Addresses From Scrapers.

Spammers crawl the Web in search of email addresses posted in clear text. Captchas provide an effective mechanism to hide your email address from Web scrapers. The idea is to require users to solve a captcha before showing your email address.(wwwcaptcha.net)

Online Polls. In November 1999, there was an online poll asking which was the best graduate school in computer science. As in the case of online polls, IP addresses of voters were recorded in order to prevent single users from voting more than once. However, students at Carnegie Mellon University (CMU) found a way to stuff the ballots using programs that voted for the university thousands of times as the university’s score started growing rapidly. The next day, students at Massachusetts Institute of Technology (MIT) wrote their own program and the poll became a contest between voting “bots.” MIT finished with 21,156 votes, Carnegie Mellon with 21,032 and every other school with less than 1,000 votes. This explains why the result of any online poll cannot be trusted until when the poll can ascertain that only humans can vote [15].

Preventing Dictionary Attacks. Captchas can also be used to prevent dictionary attacks in password systems. The idea is simple: prevent a computer from being able to iterate through the entire space of passwords by requiring it to solve a captcha after a certain number of unsuccessful logins. This is better than the classic approach of locking an account after a sequence of unsuccessful logins, since doing so allows an attacker to lock accounts at will [12].

Search Engine Bots. It is sometimes desirable to keep Web Pages unindexed to avoid them from being easily located. There is an html tag to prevent search engine bots from reading web pages. However, the tag does not guarantee that bots would not read a web page. Search engine bots, since they usually belong to large companies, respect web pages that don’t want to allow them in. However, captchas are needed in order to truly guarantee that bots won’t enter a web site [15].
Worms and Spam. Captchas also offer a plausible solution against email worms and spam. A few companies are already marketing this idea of ensuring that a user only accepts emails if they are sure there is a human behind the computer.

2.2 Problem Statement

Current CAPTCHA algorithms are plagued with certain inadequacies due to their simplicity and the easiness of the questions offered by the challenge-response system. Their timing is also flawed and this makes them more vulnerable and calls for improvement.

3. THE TM-MCRS FRAMEWORK

The intention of this research is to strengthen the application of CAPTCHA by introducing an Integrated Mathematical Mechanism with the Time Out System to a Multiple Response Challenge Captcha. The Time-Out Multiple Challenge Response System (TM-MCRS) is proposed as a mechanism where a user is expected to answer two CAPTCHA tests by providing two responses before he/she is granted access to the protected resources. The system times out the user where the user fails to get through the second challenge. The first challenge is the retype design and the second is the integrated mathematical challenge. A program will be developed based on the SHAI algorithm to generate the CAPTCHA. This will be implemented using PHP, Netbeans, and Java. The TM-MCRS is a system where the challenges are administered in a multiple of times. The first being the alphanumeric text challenge where the user is expected to retype just exactly the generated text as it appears, followed by the challenge called the integrated mathematical mechanism where the user is expected to give correct answers to a mathematical challenges for him/her to proceed. The timing out is essential in that it will reduce the chances bots have to input answers to the challenge.

The TM-MCRS framework emulates and improve the general classical CAPTCHA procedures for a single challenge system as proposed by [7] are as follow

- Computer generates a test instance.
- Test is shown to the human/bot.
- Human/bot attempts to solve the test.
- Human/bot reports suppose solution to the computer.
- Computer evaluation the submitted solution.
- Computer reports the result of evaluation to the human/bot
- Computer allows access if solution submitted is correct

Our implementation adopts the following guidelines:

Accessibility. Captchas must be accessible. It is solely based on reading text or other visual-perception tasks to prevent visually impaired users from accessing the protected resource. Such captchas may make a site incompatible with Section 508 in the United States. Any implementation of a captcha should allow blind users to get around the barrier, for example, by permitting users to opt for an audio captcha.

Image Security. Captcha images of text should be distorted randomly before being presented to the user. Many implementations of captchas use undistorted text, or text with only minor distortions. These implementations are vulnerable to simple automated attacks. (recaptcha.net)

- Script Security. Building a secure captcha code is not easy. In addition to making the images unreadable by computers, the system should ensure that there are no easy ways around it at the script level. An example of insecurities in this respect is where systems bypass the answer to the captcha in plain text as part of the web form. Captcha scripts found freely on the Web are mostly vulnerable to these forms of attacks [14]

Security Even After Wide-Spread Adoption. There are various captchas that would be insecure when a significant number of sites start using them. This attribute is inadequate as true captchas should remain secure regardless of the numbers of websites adopting them.

3.1 Design Issues

This section gives us an insight into the design of our proposed TM-MRC, which is a way of strengthening captcha system and combating cyber-crimes of different sorts. A text challenge is administered once followed by a simple mathematical test (which is the second challenge). A time out is then added to the mechanism to reduce both the time wasted on the system for each access and the time bot has to wait to deduce or solve the captcha system. This implies that once the second challenge is administered the time taken to solve the challenge starts counting expecting that the human user will have to solve the challenge at a given time and of which a bot will have same time to attempt the challenge. In a case where the bot throws the captcha words to spammer especially on some pornographic site, the time taken to solve the captcha will not meet the time that has already started counting on the site where captcha is administered. This reduces the time a bot has to solve a captcha text thereby increasing the security of resources on the site administered with captcha. This is used to screen out bots. The concept of this captcha and the previous CAPTCHAs is motivated by real-world problems faced by internet companies such as Yahoo! and AltaVista. These companies offer free email accounts, intended for use by humans. However, they found that many online vendors were using “bots” - the computer programs that would sign up for thousands of email accounts from which they could send out masses of junk emails.

3.2 Typical Captcha Image Generation

Research has shown that spam bots or users using Optical Character Recognition (OCR) software to read and capture existing texts are able to bypass CAPTCHA tests with a good level of success. The need therefore arise to present a second challenge that contains mathematical elements. This challenge will be administered to check the advances made
by bot and then the time out administered to reduce the number of trials a user or spambot requires to detect the administered challenge. The system automatically generates the challenges - a distorted text with busy background to make it unreadable for OCR - and the simple mathematical calculation all stored in the database which displays a “session has expired” message after the second failed attempt thereby reducing the number of trials a user or spam can make attempts. Below is figure 2 illustrating a typical single captcha challenge response architecture. Figure 3 depicts a Typical CAPTCHA Image Generation Implementation Framework.

Fig 2. Typical Single Response Captcha Architecture
The general algorithm for a classical CAPTCHA implementation framework described in figure 3 is as follows:
1. Computer generates a test instance;
2. Test is shown to the human/bot;
3. Human/bot attempts to solve the test;
4. Human/bot reports supposed solution to the computer;
5. Computer evaluates the submitted solution; and
6. Computer reports the result of evaluation to the human/bot and allows or blocks access to a resource based on the result.

Fig 3: A Typical Captcha Image Generation Implementation Framework
This single response captcha mechanism allows human or robot to retry as many times as possible at the time the user deems fit. This means that the bot has ample of time to send the captcha challenge to spammers who solves the captcha at their own time and the bot captures the solved challenge and then administers it to the site and gets access to whatever information it requires. This is a risk to the websites.

A Prototype of an e-mail service scenario, which is a specific instance of a single challenge system, is depicted in figure 4.

The following is obtainable in an e-mail service scenario depicted in figure 4:

1. The user makes a request such as request to open an email account or fill a form on an email account.
2. The email client requests the email server.
3. The email server asks the CAPTCHA server to generate a CAPTCHA test to verify the user: Human/ AI Assistance Program.
4. The CAPTCHA server returns a test and its answer.
5. The email server responds to email client with the CAPTCHA test.
6. The email client calls the CAPTCHA tester to deliver the challenge.
7. The test is shown to the human/AI assistance program.
8. The CAPTCHA tester collects the user’s answer.
9. The CAPTCHA tester forwards result to the email client.
10. The email client submits the solution to the email server.
11. The email server evaluates the submitted solution and passes the result to the email client.
12. The email client reports the result of evaluation to the user – human/ AI Assistance Program – and allows or blocks access to a resource based on the result.
A non-response or incorrect response to this test indicates an AI assistance program or bot. The design of the CAPTCHA server is to provide a random CAPTCHA challenge, verify the answer, and return a predicate confirming that the answer is correct. The functionality of the CAPTCHA Tester running on the user’s machine is to display a CAPTCHA challenge to the user and gets the corresponding answer.
4. ARCHITECTURE OF THE TM-MCR MECHANISM

The architecture of the proposed TM-MCR mechanism is depicted in Figure 5.

**Figure 5: Visual Representation of the Testing Procedure of the Multiple C-R System With Database considered**
5. CONCLUSION

In this work, we have proposed a Time-Out Multiple Challenge Response CAPTCHA that inculcates integrated mathematical mechanism to boost the security of CAPTCHA codes. This is to enable users overcome the limitations of the current CAPTCHA system that consist of only alphanumeric characters. The introduction of special features or characters (such as parentheses, brackets, braces and mathematical characters like Boolean characters) into the code is necessary to strengthen the CAPTCHA’s security Mechanism.

6. FUTURE WORK

Future work will implement the proposed mechanism using appropriate software tools.
REFERENCES

Survey of Opinion of Marginalized College Students About Online University Education

A.O. Itegboje
Yaba College of Technology
Yaba Lagos, Nigeria
email-jibike@yabatech.edu.ng, jibike_itegboje@yahoo.co.uk

J.A. Adedokun
Federal College of Education (T) (FCE (T))
Akoka-Yaba, Lagos, Nigeria
email-adedokun.james@yahoo.com

O. O. Olusanya
Department of Computer Science
Tai Solarin University of Education
Ijebu-Ode, Nigeria
yinka_olusanya@yahoo.com

ABSTRACT

In Nigeria, tertiary education is obtainable from the Universities, Polytechnics and Colleges of Education and preference by majority is in that order. Meanwhile, the high demand today for tertiary education renders thousands of eligible and qualified college students unadmitted and marginalized because of a limitation of space. Most students that opt for a University education end up being offered Polytechnic and College of Education admission contrary to their desire. A major alternative to this is an online University education. The purpose of this study was to determine the extent of awareness of Nigerian College students about online tertiary education and the practicability of the same in Nigeria. A Survey research design was used for the study. The population comprised students from Lagos (YABATECH and FCE (T)). Five-hundred students with 250 from each institution constituted the sampled studies. Three research questions and two hypotheses were formulated. A 25-item questionnaire was used for data collection. It was validated by two experts and its reliability estimated to be 0.83 and above. The instrument was administered by the researchers and the resulting data analyzed with f, %, X, SD and t-test statistics. The result from the study show among other things that more than 68% of Nigerian College students are aware and have tried accessing online education. It also shows that online education is practicable in Nigeria but not without some hurdles which are however surmountable. It was recommended that Government at all levels in Nigeria should invest in Information Communication Technology(ICT) in order that college students will be ICT compliant and have better access to online education.

Keywords: Marginalized, Students, College, Online, University & Education.

I. BACKGROUND OF THE STUDY

Tertiary education is viewed as the bedrock of a nation’s economic growth [3], Nigeria, a developing nation, with a populace of about 140 million, is only able to give admission annually to less than 20% of those seeking admission into its 104 universities (JAMB, Annual Report 1996 – 2000). Tertiary education is obtainable from three types of institutions, universities, polytechnics and colleges of education. Generally, the order of public and individual’s rating, acceptance, embrace and preference for these institutions and their certificates is first for the universities followed by the polytechnic and last for the colleges of education.

The reasons for this are not far-fetched. It includes the fact that some see it as a status symbol while to some others; it is a major determinant of the level of intelligence of an individual. To these set of people it is those that are most intelligent that are absorbed into the university system [6]. More important, universities are believed to be institutions where professionals (Engineers, Medical doctors, Bankers, Accountants) are created. On the contrary, to make matters worse, the situation employers of labour both private and public give preference to university degrees over diplomas and certificates obtained from the other tertiary institutions. In essence, employment is often not by merit but tied to the nature of the institution attended.[2]
This is not only limited to the labour market, but admission for postgraduate studies also show preference for university degrees over certificates from polytechnics and colleges of education. Polytechnics are often rated as glorified Technical Colleges where technicians who are meant to serve as support staff to engineers are groomed, while Colleges of Education are viewed to be exclusively for teachers in training. Universities are given the exclusive right to award degrees while polytechnics and colleges of education are meant to award diplomas or certificates. Based on the aforementioned reasons, Nigeria’s college students show high preference for the University Education and only opt for other tertiary institutions where all effort to secure University admission has failed. Meanwhile, Nigeria has relatively few public and private universities. The number of applicants every year is always overwhelming to the extent that most of the candidates who are able to score the cut-off marks are often not admitted. This is exemplified by the 2010 Unified Tertiary Matriculation Examination (UTME) where as many as 867,000 candidates were able to meet the 180 cut-off mark. According to Dibu Ojerinde, who sits on the Registrar of Joint Admission Matriculation Board, which is the body responsible for conducting admission for universities, polytechnics and colleges of education, “As many as 340,000 of these candidates will not be admitted into any of the nation’s tertiary institution this year”. “The nation’s tertiary institutions could not absorb all the candidates because of ‘Limited space’ “. He further described admission into tertiary institution this year as a “survival of the fittest” and that the space available for prospective students was about 527,000. The Registrar also disclosed that most of the candidates did not choose many of the privately-owned universities because of high tuition fee. In addition, Dibu Ojerinde has suggested “unless tertiary institutions seek efficient methods to progress, many prospective students will wait endlessly for admission into college. [9]

In essence, nearly 340,000 students that met the cut-off scores but without admission have become marginalized, while the situation becomes worse each year because other sets of qualified students add on to the waiting list. To circumvent this situation, alternative must be sought. One method which has proven to be successful at other institutions of learning today is the online tertiary education. According to experts, online courses are those in which at least 80% of the course content is delivered online. Face-to-face instruction, on the other hand includes courses in which zero to 29% of the content is delivered online. This category includes both traditional and web-facilitated courses. The remaining alternative, which is the blended instruction type, is defined as having between 30% and 80% of the course content delivered online.[5][10]

Another major tertiary education alternative open to college students is a form of face-to-face education in other nations. Some nations have embraced this and it has worked for them. For instance, as of 2009, data on the international studentship in American Universities and India had 861,000 students enrolled in American Universities. China had about 400,000 to 500,000, while Japan, Korea and the continent of Africa together had less than 28,000 enrollees. This is because Africans, especially Nigerians have misplaced their priorities for other things, and the only solution is for their priorities to be reordered; more so as no nation can thrive on ignorance.

The monetary implication of being an international student which would have been a way out for marginalized students may not be affordable for the majority of Nigerian College students because of the level of poverty in the nation. Nigeria's per capita income stands at $2,748, while Nigerians that live on $1.25 per day make 29.6 per cent of the population, while those that live on $2 are 83.9 per cent (The World Bank 2010 World Development Report). Online education therefore appears to be an option to be considered as it affords one the opportunity of work and study simultaneously. One also has the opportunity to earn an international degree while residing in a location that is convenient for one. It also affords the opportunity of studying at ones convenience. This study is borne out of the highlights above.[8]

1.1 Definition of Terms

- University- A four or five year straight bachelors degree awarding institution
- Polytechnic- A five year program broken into three segments, first two year completion leads to award of National diploma, next one year is used for Industrial Attachment, followed by another 2-year program leading to award of Higher National Diploma. Graduates of this program are supposed to be rated as equivalent to University degree but that is not the reality
- Colleges of Education- A straight 3 year program leading to the award of National Certificate of Education.

1.2 Purpose of the Study

The order of preference for tertiary education among eligible and qualified Nigerian students is first for the University system and as a last resort, the Polytechnics and Colleges of Education. The reasons for this are not far-fetched. First, is the fact that the masses see University education as being superior, and second, the employers of labour discriminate against Polytechnic and Colleges of Education graduates in preference for university graduates. Such is to attribute for what marginalizes Polytechnic and Colleges of Education graduates. Presently, universities from all over the world are becoming accessible online. The purpose of this study is to find out:

1. The extent of awareness of online tertiary education among Nigerian students.
2. Whether online education can be effectively practiced in Nigeria.
3. What will happen to Polytechnic and Colleges of Education if online education becomes easily accessible and affordable?
1.3 Research Questions
1. What is the extent of awareness of online university education among Nigerian students?
2. Is online university education practicable in Nigeria?
3. Will awareness, accessibility, and affordability of online tertiary education reduce the patronage of Nigerian Polytechnic and Colleges of Education?

1.4 Research Hypotheses

$H_{01}$: There is no significant difference between opinion of students about awareness and accessibility of online education among the Polytechnics and Colleges of Education students.

$H_{02}$: There is no significant difference between the awareness and accessibility of Nigerian male and female students to online education.

2. RESEARCH DESIGN

A survey research design will be adopted for this study. In addition, a representative sample of the student population will be selected and results from the study will be used to generalize the rest of the population.

2.1 Population, Sample and Sampling Technique

The tertiary education system in Nigeria comprises the Universities, the Polytechnics and Colleges of Education. The population for this study consists of all the students in one Federal Polytechnic and one College of Education in Lagos State. They are Yaba College of Technology (YABATECH) and Federal College of Education Technical (FCE (T)). The choice of these institutions is based on the fact that the two offer mostly technical, vocational and engineering courses. It is also based on the fact that the majority of the students in these institutions opted for it because they failed to secure admission to Universities which they would have being their first choice of preference.

A stratified random sample consisting of 500 students was taken to conduct the study in which 250 students were selected from each of the two institutions. The two institutions were first stratified into Polytechnic and College of Education. From each stratum, representative sample were selected randomly.

2.2 Instrument for Data Collection

A twenty-five item questionnaire “Survey of Opinion of Students about Online Education (SOSOE)” was used to prove the hypothesis. The survey had two major sections labelled A and B. Section A consisted of items that queried students for bio-data (age, gender, institution, etc.) Section B comprised 22 structured items which are designed to elicit information about students’ awareness, access, affordability and preference for online University education. They were also meant to provide answers to the research questions and hypotheses formulated to guide the study.

2.3 Validation of the Questionnaire

The questionnaire was designed by the researchers based on the purpose of the study, literature review, research questions, and hypothesis formulated. It was however given to two colleagues who are co-lecturers and vast in research and questionnaire design. They helped to correct all forms of errors. They reframed some of the items, expunged some and replaced them with others. They also suggested other areas of modification which were affected by the researchers. All these efforts helped to ensure that the instrument had high content and construct validity.

2.4 Reliability of the Instrument

To ensure its consistency, the reliability of the questionnaire was estimated before being used for data collection. This was done through the administration of 40 copies of the validated questionnaire to 40 respondent students who are not part of the main study, with 20 selected from each of the two institutions. The data generated was divided into two equal halves with the use of split-half method of estimating reliability and analyzed statistically with Pearson’s Product Moment Correlation. The result gave a correlation (reliability) coefficient of 0.831 and above. This helped to ensure that the questionnaire was highly reliable.

3. METHOD OF DATA COLLECTION

The questionnaire was personally administered by the researchers. The purpose and the eventual gain that would accrue to students after successful completion of the study were explained to students. Next, the students were encouraged to complete the questionnaire properly while the result of not completing the questionnaire properly was also explained. Due to close monitoring, the return rate was very high.

3.1 Method of data Analysis

The data generated with the questionnaire were analyzed with both descriptive and inferential statistical tools which include:

- Frequency count (ff)
- Percentage (%)
- Correlation (r)
- Mean (X)
- Standard deviation and
- T-test
3.2 Data Analysis and Presentation

Table 1: Opinion of Nigerian Students about their Awareness and Effectiveness of Online University Education

<table>
<thead>
<tr>
<th>S/No.</th>
<th>Awareness of online university education</th>
<th>Agree</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>I have idea of what online university education is</td>
<td>337</td>
<td>68.92</td>
</tr>
<tr>
<td>2.</td>
<td>I have sourced for online education before</td>
<td>281</td>
<td>57.46</td>
</tr>
<tr>
<td>3.</td>
<td>I prefer online university education to regular polytechnic/college of education</td>
<td>256</td>
<td>52.35</td>
</tr>
<tr>
<td>4.</td>
<td>Online university education is common with postgraduate programs</td>
<td>388</td>
<td>79.35</td>
</tr>
<tr>
<td>5.</td>
<td>Skills that will be acquired in face to face school may not be different from online education.</td>
<td>240</td>
<td>49.08</td>
</tr>
<tr>
<td>6.</td>
<td>There is discrimination between online degrees and degrees obtained from face to face institutions.</td>
<td>308</td>
<td>62.99</td>
</tr>
</tbody>
</table>

Table 1 shows the data analyzed about the opinion of Nigeria Polytechnics and Colleges of Education students about online University Education. It reveals that about 68.9% of the 500 students used for the study have the knowledge on online University Education. It also shows that more than 57% have tried online education, about 52% prefer online university education, 49% belief that skill acquired from online education is not different from that of face to face classroom situation. It reveals further that about 63% are of the opinion that online degrees are discriminated against.

Table 2: Opinion of students about practicability of online education in Nigeria

<table>
<thead>
<tr>
<th>S/No.</th>
<th>Practicability of online education in Nigeria</th>
<th>Agree</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.</td>
<td>Online university education is possible in Nigeria.</td>
<td>317</td>
<td>64.8</td>
</tr>
<tr>
<td>8.</td>
<td>Nigeria governments do not give much recognition to online education</td>
<td>422</td>
<td>86.3</td>
</tr>
<tr>
<td>9.</td>
<td>The present state of internet infrastructure will not make online education practicable.</td>
<td>285</td>
<td>58.3</td>
</tr>
<tr>
<td>10.</td>
<td>The cost of computer and internet access will not make online education workable.</td>
<td>328</td>
<td>67.1</td>
</tr>
<tr>
<td>11.</td>
<td>Practicability of online education is tied to knowledge of ICT among the student.</td>
<td>401</td>
<td>82.0</td>
</tr>
<tr>
<td>12.</td>
<td>Online education is not easily affordable because it is costly.</td>
<td>415</td>
<td>84.9</td>
</tr>
<tr>
<td>13.</td>
<td>Lack of constant electricity will not make online education practicable in Nigeria.</td>
<td>372</td>
<td>76.1</td>
</tr>
</tbody>
</table>

The analyzed data in table 2 reveals the opinion of Polytechnic and College of Education students about the practicability of online education in Nigeria. Among other things, the data shows that about 65% observed that online education is possible in Nigeria, 86% noted that Nigerian government do not give much recognition to online education, 82% believe that online education is tied to the students' knowledge of ICT while 85% are of the opinion that it is very costly and unaffordable.
Table 3: Opinion of students about factors in favour of online university education

<table>
<thead>
<tr>
<th>S/No</th>
<th>Factors in favour of online education</th>
<th>Agree</th>
<th>Mean</th>
<th>F</th>
<th>%</th>
<th>X</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>14.</td>
<td>If online university education is easily accessible, I will prefer it to Polytechnic/College of Education</td>
<td>364</td>
<td>74.4</td>
<td>3.45</td>
<td>1.18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15.</td>
<td>Nigerian students resort to Poly/COE because they cannot gain admission into University</td>
<td>352</td>
<td>72.0</td>
<td>2.97</td>
<td>1.18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16.</td>
<td>The labour market favours University degree over Poly/COE degrees.</td>
<td>382</td>
<td>78.1</td>
<td>3.02</td>
<td>1.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17.</td>
<td>There is general increase demand for online education among students.</td>
<td>315</td>
<td>64.4</td>
<td>2.56</td>
<td>3.11</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

According to the analyzed data in table 4, about 61.6% of the respondents noted that they are proficient at using computer. Only about 43% of the students indicated that they know how to use internet and the resources available on it while majority (98.2%) are aware that online education is not possible without the knowledge of ICT.

Table 5: t-test of difference between awareness and accessibility of online education among polytechnics and colleges of education students

<table>
<thead>
<tr>
<th>Institution</th>
<th>N</th>
<th>SD</th>
<th>X</th>
<th>df</th>
<th>t-cal.</th>
<th>t-cri.</th>
<th>Remak</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polytechnic</td>
<td>24</td>
<td>7</td>
<td>2.1</td>
<td>5</td>
<td>0.0</td>
<td>1.2</td>
<td>H01</td>
</tr>
<tr>
<td>College of Education</td>
<td>24</td>
<td>2</td>
<td>1.9</td>
<td>4</td>
<td>3.3</td>
<td>3.0</td>
<td>Accepted</td>
</tr>
</tbody>
</table>

Table 5 reveals t-test of difference between the opinion of 247 Polytechnic and 242 College of Education students about their awareness and accessibility to online education. It shows a calculated t-test value of 1.20 and a critical value of 2.00 at 0.05 level of significance and 487 degrees of freedom. This implies there is no significant difference between the opinions of the students; hence, the null hypothesis was accepted.

Table 6: t-test of Difference Between Opinion of Male and Female Students to Online Education

<table>
<thead>
<tr>
<th>Gender</th>
<th>N</th>
<th>SD</th>
<th>X</th>
<th>df</th>
<th>t-cal.</th>
<th>t-cri.</th>
<th>Remak</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>22</td>
<td>4</td>
<td>3.0</td>
<td>7</td>
<td>3.3</td>
<td>1.1</td>
<td>Accept</td>
</tr>
<tr>
<td>Female</td>
<td>26</td>
<td>5</td>
<td>1.4</td>
<td>7</td>
<td>3.1</td>
<td>1.6</td>
<td>Accept</td>
</tr>
</tbody>
</table>

Table 6 reveals a t-test calculated value of 1.16 and a corresponding critical value of 2.00. The fact that the calculated value is lower than the critical value implies that there is no significant difference between the opinion of male and female students about their awareness and access to online education. Hence, the second hypothesis was also accepted.

5. SUMMARY OF FINDINGS

The major findings from this study include:
1. More than 68% of Nigerian college students have an idea of online university education while about 58% have attempted sourcing for it.
2. About 52% of the Polytechnic and College of Education students prefer online university education while as much as 79% believe that online education is common and much more accessible for postgraduate programs.
3. More than 62% of the respondents indicated that there is discrimination between online degrees and those obtained from face-to-face institutions. Only 49% believe the skills obtained from online education are not different from those obtained from face-to-face institutions.

4. About 65% respondents believe that online university education is practicable in Nigeria while about 86% noted that Nigerian governments do not give it recognition.

5. Factors against the practicability of online education in Nigeria include the state of internet infrastructure, cost of computers and internet access, level of ICT know-how among students and lack of constant electricity supply. These are supported by about 58%, 67%, 82% and 76% of respondents respectively.

6. Factors in favour of online university education in Nigeria include difficulty of gaining admission to Nigerian universities and discrimination of the labour market towards polytechnic and colleges of education certificates. These are supported by 72% and 78% of the respondents respectively.

7. On the knowledge of ICT, the study shows that only 61.6% of the students are proficient in the usage of computer and 43.2% in the usage of internet and its resources.

8. There is no significant difference between the opinion of polytechnic and college of education students, and also the opinion of male and female students about their awareness and practicability of online university education in Nigeria.

6. DISCUSSION AND IMPLICATION OF FINDINGS

Presently in Nigeria, admission into the higher institutions is “a survival of the fittest” because of the limited space in the universities. In addition, the masses generally parents and even their children believe and even show significant preference for University education over Polytechnic and Colleges of Education. This preference is a status symbol attached to socio-economic status among Nigerian society. Added to this is that the employers of labour and the labour market generally do not show any sign of understanding as they even discriminate against other certificates other than those obtained from the universities.

To circumvent this situation, an alternative being urgently proposed is online university education. One of the first findings of this study reveals that about 52.4% of Nigerian Polytechnic and College of education students indicated preference for online university education to their present institutions of learning, while another 58% of the respondents observed that they have attempted sourcing for online university education before they resolved to opting for Polytechnic or College of Education because they had acquire further education. Furthermore, this study revealed about 49% of the respondents believe that online education would be as effective as the face-to-face form of education.

This is corroborated by the research works of Clark (1983) and Russel (1999), where they observed that a great deal of evidence exist showing no significant differences should be expected regarding the effectiveness of a well-designed face-to-face learning format. The reason for this is that as the popularity of the Internet grows so does the potential for online learning (Muilenburg & Berge, 2005). The only significant difference that may be expected is how students perceive their online experiences during learning. To the extent that the students’ perceptions are negative regarding their past, present, or future online learning experiences, the students’ perception may contribute to such outcomes as higher dropout rates (Carr, 2000), low motivation of students to learn (Maltby & Whittle, 2000), and lower student satisfaction with the learning experience (Kenny, 2003).

In answer to the question of whether online education is practicable in Nigeria, the study found out that the opinion of the desperate and marginalized students is in the affirmative. Meanwhile about 85% said that it would be cost intensive and about 82% observed that its practicability is going to be based on the students’ level of ICT know-how; on the other hand, 76% respondents said one of its major constraints is a lack of consistency of a power supply in Nigeria. This study also sought to establish whether awareness, accessibility and affordability of online tertiary education will reduce the patronage of students to Nigerian polytechnics and Colleges of Education. More than half of the respondents (52.35%) agreed to this. This implies that the present level of patronage of these institutions is at the mercy of awareness, accessibility and especially affordability of online education.

This is supported by Allan and Seaman (2010) in their executive summary of an article titled “Learning on Demand” where they stated that online enrolments have continued to grow at rates far in excess of the total higher education students’ population, with the most recent data demonstrating no signs of slowing.

The study also found out that about 64% of the students indicated that there is an increased demand for online education among Nigerian students. This is supported by the first and second hypothesis which revealed that there is no significant difference between the opinion of Polytechnic and College of Education students, and among male and female students about their awareness of online university education and their increased desire for online university education. This appears peculiar to developing nations who want to cross over to developed nations because of the developments in those areas, under the guise of search for university education. This is because the work of Allan and Seaman (2010) further reported that in developed nations, especially United States, over one-half (54%) of institutions report that the economic downturn has increased demand for existing face-to-face courses. They further observed that the economic impact has been greatest on demand for online courses.
7. CONCLUSION AND RECOMMENDATIONS

Online education abounds as a major alternative to university education in Nigeria which has been over-subscribed by eligible and qualified candidates. This has rendered some college students to be marginalized and so resulted to seeking admission to Polytechnics and Colleges of Education against their desires and aspiration. The first purpose of this study was to find out the extent of awareness of college students about online university education. Based on the findings, it can be concluded that the college students are not only aware of it but are aggressively pursuing it as an alternative to Polytechnic and College of Education. In terms of the practicability of online education in Nigeria, we conclude based on the findings that it can be effectively practiced, although it may meet some challenges, such as poor electricity which can be overcome as today’s laptop come with long power back up. Also low level of ICT compliance of the nation may abe a hinderance. However, according to World Bank report there has been appreciable increase in IT infrastructure and services that can support implementation of online education (see Table 7).

Table 7- Facilities supporting Online Education

<table>
<thead>
<tr>
<th>Internet Penetration-Users</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
</tr>
</thead>
<tbody>
<tr>
<td>Secured internet access per 1 million people</td>
<td>0.17</td>
<td>0.25</td>
<td>0.49</td>
<td>0.79</td>
<td>0.81</td>
</tr>
</tbody>
</table>

| ICT goods import | 6.9% | 4.5% | 10.2% |


We also conclude that the present level of patronage of Nigerian Polytechnics and Colleges of Education is seriously at the mercy of awareness, accessibility and affordability of online university education.

We therefore recommend that:

1. All tiers of government in Nigeria should take ICT very seriously by investing in it. This will allow more college students to become ICT compliant and have better access to online university education;
2. Government should rise to the occasion of stopping every form of discrimination between polytechnic, College of Education and University certificates, either for employment purpose or for further studies.
3. The government should also ensure the same certificates are obtainable from all higher institutions, either by reviewing and increasing the curricular contents proportionally and or by increasing the number of years for the respective programs.
REFERENCES


Load Balancing Adaption of Some Evolutionary Algorithms In Cloud Computing Environment

A.P. Priyadarshini, R. Sharmistha & K.P. Prasant
School of Computer Engineering
KIIT University
Bhubaneswar, India
priyadarshinikiit@gmail.com, sroyfcs@kiit.ac.in, ,patnaikprasantfcs@kiit.ac.in

ABSTRACT

This paper aims to conduct an experiment on using some popular algorithms say Particle Swarm Optimization Load Balancing Algorithm (PSO), Cat Swarm Optimization Load Balancing Algorithm (CSO) and Genetic Load Balancing Algorithm (GA) in dynamic cloud computing environment in order to achieve low overheads during load balancing.

Keywords: Response Time, Cost, Load Balancing, Adaption, Evolutionary Algorithms & Cloud Computing Environment

African Journal of Computing & ICT Reference Format:

1. INTRODUCTION

In Cloud Computing Environment the whole re-distribution of load takes place in various environments i.e static and dynamic. Static state refers to the current form or present situation where as dynamic rebels runtime statistics which focuses both on past and present. Dynamic environment is the user friendly state as it lies in the flexibility of user requirements. The whole cloud environment is balanced in centralized, distributed and hierarchical designs. Among the three approaches, huge memory management and maximum nodes structuring is done in distributed balancing environment.

Load balancing is essential in order to decrease load overhead, reduce bottleneck and delays, and improve tolerance. Balancing of load focuses on managing nodes, managing resources, grids, clusters, networks, etc. Centralized balancing rebels the global synchronization strategy in one node where as distributed balancing defines the collaboration of multiple nodes with individual processing's. Centralized balancing takes less time in analyzing different resources and in processing's. In distributed balancing scenario, if one node fails than it doesn't affect the rest nodes and no overloading takes place.

Distributed balancing approach is scalable, decentralized in nature and even support Grid systems. Here the whole scheduling is done by requesters independently. Some of the examples of distributed balancing algorithms are Ant Colony optimization algorithm, Map Reduce, Particle Swarm Optimization Load Balancing Algorithm, Cat Swarm Optimization Load Balancing Algorithm, Active Clustering algorithms and many more. Where as in the other way Round Robin, Genetic Load Balancing Algorithm, Max Min and Min Min are some of the centralized balancing algorithms. Hierarchical balancing in total is modeled as tree data structure where each node of the tree is level headed. Different levels of the cloud in load balancing is included in hierarchical balancing. Map Reduce algorithm is one of the exemplar of hierarchical balancing.

The remaining part of the paper is organized as follows: Section 2 focuses on the parallel work on load balancing and section 3 states the proposed Cat Swarm Optimization Load Balancing Algorithm technique in Cloud Computing Environment. Section 4 emphasis the relative survey and the experimental outcomes obtained after the study through Cloud Sim and Cloud Analyst, Section 5 overviews the conclusion part of the paper.
2. RELATED WORKS

The load balancing in Cloud Computing Environment have been one of the newly emerging concept in IT world. Many clients from various regions of the world demand for good quality and faster service. In order to satisfy the clients demand and for proper resources allocation, load balancing is performed using effective algorithms. The following papers describe some of the related works done in Cloud Computing Environment using load balancing algorithms:

Shu-Chuan Chu and Pei-Wei Tsai in paper [7] include an improved Genetic Load Balancing Algorithm for function optimization which concludes with effective simulation results than other traditional algorithms.

P.J. Angeline in paper [8] describes briefly the Particle Swarm Optimization Load Balancing Algorithm with analysis of its advantages and disadvantages in complex optimization problems. Particle Swarm Optimization Load Balancing Algorithm in coordinate system, non-coordinate system has also been described with applications.

Mansouri and Mohammad Teshnehlab in paper [2] introduced an improved Cat Swarm Optimization Load Balancing Algorithm which on comparison with pure Cat Swarm Optimization Load Balancing Algorithm. The whole comparison states that Cat Swarm Optimization Load Balancing Algorithm clustering algorithm is better than other two algorithms.

Bhathiya Wickremasingle, R.N. Calheiros and Rajkumar Buyya in paper [9] have introduced a tool named Cloud Analyst with large scale applications using various virtual machines, data centers, service brokers, load balancing algorithms and many other services with low Cost.

Mayanka Katyal and Atul Mishra in paper [10] presents different load balancing algorithms in Cloud Computing Environment according to service level agreement. This paper aims in achieving less Response Time with maximum resource utilization.

3. CAT SWARM OPTIMIZATION LOAD BALANCING ALGORITHM IN CLOUD COMPUTING ENVIRONMENT

Load balancing in Cloud Computing Environment can be successfully achieved by reducing the Response Time with low Cost and widely utilizing the resources. With the wide development of web, processing time plays an important role in balancing the loads. The huge demand and use of web has led to building large number of processors, workstations, PCs, etc together in organizing a distributed computing environment. In order to improvise the processing speed and efficiency of the system, division of workloads to every individual node is done. Various algorithms based for distributed computing environment are used, which would lead in balancing the loads. Cat Swarm Optimization Load Balancing Algorithm is one of distributed balancing algorithm considered for load balancing. Cat Swarm Optimization Load Balancing Algorithm Load is one of the newly emerged algorithm in optimization field of computational intelligence of natural evolution. Cat Swarm Optimization Load Balancing Algorithm has been turn out as one of the Cost effective with low Response Time stochastic meta-heuristic optimization algorithm. In this paper, Cat Swarm Optimization Load Balancing Algorithm has been proposed in Cloud Computing Environment as a distributed computing load balancing technique to find the best solution in less time. Below the proposed approach is described in the next section.
3.1. Proposed Cat Swarm Optimization Load Balancing Algorithm in Cloud Computing Environment

Cat Swarm Optimization Load Balancing Algorithm is an emerged new interesting optimization algorithm. In this algorithm, the awareness of cats is considered. Cat Swarm Optimization Load Balancing Algorithm is considered as it provides faster and much better solutions for various problems in some fields. The behavioral state of the cat is divided into two forms i.e : Seeking process and Tracing process.

1. Seeking process : This state rebels the resting period of the cat and even the alert for next move.
2. Tracing process : It resembles the target searching stage.

The movement of the cats are enclosed in a D-dimensional space considering four different parameters i.e position , velocity, fitness function and flag. The flag is used in order to identify the different processes. In optimization search problems, cat is considered because this breed spends more time in resting rather in tracing. As the tracing process takes very less time and more energy is utilized ,so in this algorithm a mixture ratio is considered. Mixture ratio combines both the processes i.e seeking process and tracing process .The position vector (Pi ) equation is represented below by 1 :

\[ P_i = \frac{(FVi - FVb)}{(FV_{max} - FV_{min})} \]  

(1)

Similarly the attributes of velocity vector (V_{r,d} ) can be represented by 2.

\[ V_{r,d} = V_{r,d} + r1 \cdot c1 \cdot (p_{best,d} - P_{r,d}), \quad d = 1, 2, ..., M \]  

(2)

\[ P_{r,d} = P_{r,d} + V_{r,d} \]  

(3)

whereas ,P represents as the position of agent i, V represents as the velocity of agent i, pbest depicts as the most optimist state of agents, FV as Fitness value of all agents, and i ranges from 0 to j letting j equals to (SMP -1).’ j ’represents as the overall copies of ‘n’ number of cats in the memory pool.

SMP represents as the seeking memory pool (each cat seeking memory pool), the p_{best,d} as the position of the cat carrying the best fitness value and P_{r,d} represents the flag cat position with r ranging with a random value of [0,1]. For example, by considering 200 iterations we gradually found the best Cost of the cat using Cat Swarm Optimization Load Balancing Algorithm with Roulette Wheel Selection process as depicted in Fig.2.

![Fig. 1: The Seeking and Tracing Process](image)

![Fig. 2: Analysis of identifying best Cost cat in Cat Swarm Optimization Load Balancing Algorithm](image)

The proposed Cat Swarm Optimization Load Balancing Algorithm over all process is given below in Fig.3:

Step 1 : Consider n number of cats.

Step 2 : Then the cats are sprinkled randomly in D-dimensional searching space with maximum velocity value each.

Step 3 : Randomly some of the cats are picked and are entered to tracing process as per the mixture ratio and other left out cats are entered into seeking process.

Step 4 : Next each cat is evaluated using fitness function , and the best cat is considered naming pbest.

Step 5 : The cats are moved considering their flags, if the cat is in seeking state then it is moved to seeking process otherwise moved to tracing process.
Step 6: Again re-choose group of cats and move them to tracing process as per mixing ratio, and others to seeking process.

Step 7: If the termination principle gets satisfy then the program terminates otherwise Step 4 to Step 6 repeats.

4. SIMULATION RESULTS AND ANALYSIS

The proposed Cat Swarm Optimization Load Balancing Algorithm is simulated using few simulator tools like Cloud Sim and Cloud Analyst.

4.1 Simulation Toolkit Cloud Analyst

The whole Cloud Analyst architecture is built on Cloud Sim and graphical user interface. Cloud Analyst configure the whole simulation using functionalities of Cloud Sim and does simulation to get accurate results. Based on the load balancing techniques the tool evaluates the results and represents them in graphical structure. Fig. 3. represents the whole architecture of Cloud Analyst with region boundaries and simulation operations.

Fig. 3. Flowchart of Cat Swarm Optimization Load Balancing Algorithm

Fig. 4. Snapshot of the simulator "Cloud Analyst"

The whole world is divided hypothetically into 6 regions or "User bases" in Cloud Analyst simulator. These six User bases are framed in a group of users in one time zone. Table 1 represents the online registered users during peak and off-peak hours. The six regions are North America, South America, Europe, Asia, Africa and Oceania with six User bases respectively. Each data centre arranges a set of virtual machines having 4GB of RAM, storage of 100 GB, each machine carrying 4 CPUs and a power of 10K MIPS. The overall Response Time and Cost with Genetic Load Balancing Algorithm is obtained by using Cloud Analyst simulator tool where Response Time is defined in total as the arrival and waiting time of each job.
### Table 1: User base Configuration of simulation environment

<table>
<thead>
<tr>
<th>Sl.No</th>
<th>User Base</th>
<th>Region</th>
<th>Online Users in Peak</th>
<th>Online Users in Off-Peak</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>UB0</td>
<td>North America</td>
<td>5,20,000</td>
<td>30,000</td>
</tr>
<tr>
<td>2</td>
<td>UB1</td>
<td>South America</td>
<td>2,1000</td>
<td>1,000,000</td>
</tr>
<tr>
<td>3</td>
<td>UB2</td>
<td>Europe</td>
<td>3,70,000</td>
<td>5000</td>
</tr>
<tr>
<td>4</td>
<td>UB3</td>
<td>Asia</td>
<td>8,00,000</td>
<td>35,000</td>
</tr>
<tr>
<td>5</td>
<td>UB4</td>
<td>Africa</td>
<td>6,30,000</td>
<td>1,50,000</td>
</tr>
<tr>
<td>6</td>
<td>UB5</td>
<td>Oceania</td>
<td>3,50,000</td>
<td>23,000</td>
</tr>
</tbody>
</table>

### 4.2 Simulation Setup and Analysis

The whole scenario of load balancing techniques is done using Cloud Analyst simulator by considering six data centers with VMs ranging as 25, 50, and 75 for six regions of the world. Table 2 depicts with calculated average Response Time with Cost Details for Genetic Load Balancing Algorithm, Particle Swarm Optimization Load Balancing Algorithm, Cat Swarm Optimization Load Balancing Algorithm. The overall performance survey graph is depicted in Fig. 3. Subsequently, one, two, three, four, five, and six data centers with 25, 50, and 75 VMs together for each cloud configuration as mentioned in Table 2, 3, 4, 5, 6, and 7.

The simulation scenario and the overall Response Time obtained is represented graphically in the below Fig. 5, 6, 7, 8, 9, and 10. Cat Swarm Optimization Load Balancing Algorithm includes some basic operations such as initialization, fitness value calculation, seeking process, tracing process, mixture ratio, and optimal solution. The complexity of this algorithm includes time and space complexity analysis. So, by combining two processes into the algorithm, we determine the mixture ratio which rebels the rate of combining of seeking and tracing process. For example, if the number of initialized cats is 30 and the mixture ratio parameter is 0.7, then 30*0.7 = 21 number of cats are send to seeking process and remaining 9 cats to tracing process. Such way time and space complexity is determined in order to obtain an optimal solution.

### Table 2: Overall calculated Response Time of one Datacenter

<table>
<thead>
<tr>
<th>Sl.No</th>
<th>Configuratio n</th>
<th>Datacenter with VM</th>
<th>Response Time of Cat Swarm Optimization Load Balancing Algorithm</th>
<th>Response Time of Particle Swarm Optimization Load Balancing Algorithm</th>
<th>Response Time of Genetic Load Balancing Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C1</td>
<td>DC1 with 25 VMs</td>
<td>319.02</td>
<td>319.03</td>
<td>320.10</td>
</tr>
<tr>
<td>2</td>
<td>C2</td>
<td>DC1 with 50 VMs</td>
<td>318.97</td>
<td>319.01</td>
<td>319.41</td>
</tr>
<tr>
<td>3</td>
<td>C3</td>
<td>DC1 with 75 VMs</td>
<td>233.00</td>
<td>319.35</td>
<td>319.67</td>
</tr>
</tbody>
</table>

![Fig. 5. Performance study of proposed Cat Swarm Optimization Load Balancing Algorithm with Particle Swarm Optimization Load Balancing Algorithm and Genetic Load Balancing Algorithm using one data center](image-url)
### Table 3: Overall calculated Response Time of two datacenters

<table>
<thead>
<tr>
<th>Sl.No</th>
<th>Configuration</th>
<th>Datacenter with VM</th>
<th>Response Time of Cat Swarm Optimization Load Balancing Algorithm</th>
<th>Response Time of Particle Swarm Optimization Load Balancing Algorithm</th>
<th>Response Time of Genetic Load Balancing Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C1</td>
<td>DC1,DC2 with 25 VMs each</td>
<td>350.76</td>
<td>355.33</td>
<td>361.26</td>
</tr>
<tr>
<td>2</td>
<td>C2</td>
<td>DC1,DC2 with 50 VMs each</td>
<td>345.73</td>
<td>350.15</td>
<td>357.49</td>
</tr>
<tr>
<td>3</td>
<td>C3</td>
<td>DC1,DC2 with 75 VMs each</td>
<td>345.31</td>
<td>349.73</td>
<td>354.77</td>
</tr>
<tr>
<td>4</td>
<td>C4</td>
<td>DC1,DC2 with 25,50 VMs each</td>
<td>340.59</td>
<td>346.72</td>
<td>352.92</td>
</tr>
<tr>
<td>5</td>
<td>C5</td>
<td>DC1,DC2 with 25,75 VMs each</td>
<td>341.56</td>
<td>347.22</td>
<td>354.44</td>
</tr>
<tr>
<td>6</td>
<td>C6</td>
<td>DC1,DC2 with 75,50 VMs each</td>
<td>342.02</td>
<td>346.03</td>
<td>351.60</td>
</tr>
</tbody>
</table>

### Table 4: Overall calculated Response Time of three datacenters

<table>
<thead>
<tr>
<th>Configuration</th>
<th>Datacenter with VM</th>
<th>Response Time of Cat Swarm Optimization Load Balancing Algorithm</th>
<th>Response Time of Particle Swarm Optimization Load Balancing Algorithm</th>
<th>Response Time of Genetic Load Balancing Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>DC1,DC2,DC3 with 25 VMs each</td>
<td>340.33</td>
<td>345.83</td>
<td>353.36</td>
</tr>
<tr>
<td>C2</td>
<td>DC1,DC2,DC3 with 50 VMs each</td>
<td>340.20</td>
<td>344.25</td>
<td>353.52</td>
</tr>
<tr>
<td>C3</td>
<td>DC1,DC2,DC3 with 75 VMs each</td>
<td>336.03</td>
<td>340.73</td>
<td>346.18</td>
</tr>
<tr>
<td>C4</td>
<td>DC1,DC2,DC3 with 25,50,75 VMs each</td>
<td>340.00</td>
<td>342.05</td>
<td>345.21</td>
</tr>
</tbody>
</table>

### Table 5: Overall calculated Response Time of four datacenters

<table>
<thead>
<tr>
<th>S/N</th>
<th>Configuration</th>
<th>Datacenter with VM</th>
<th>Response Time of Cat Swarm Optimization Load Balancing Algorithm</th>
<th>Response Time of Particle Swarm Optimization Load Balancing Algorithm</th>
<th>Response Time of Genetic Load Balancing Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C1</td>
<td>DC1,DC2,DC3,DC4 with 25 VMs each</td>
<td>338.84</td>
<td>344.35</td>
<td>349.35</td>
</tr>
<tr>
<td>2</td>
<td>C2</td>
<td>DC1,DC2,DC3,DC4 with 50 VMs each</td>
<td>335.44</td>
<td>340.70</td>
<td>346.93</td>
</tr>
<tr>
<td>3</td>
<td>C3</td>
<td>DC1,DC2,DC3,DC4 with 75 VMs each</td>
<td>330.66</td>
<td>336.46</td>
<td>341.08</td>
</tr>
<tr>
<td>4</td>
<td>C4</td>
<td>DC1,DC2,DC3,DC4 with 25,50,75 VMs each</td>
<td>327.88</td>
<td>334.21</td>
<td>341</td>
</tr>
</tbody>
</table>

Fig. 6. Performance study of proposed Cat Swarm Optimization Load Balancing Algorithm with Particle Swarm Optimization Load Balancing Algorithm and Genetic Load Balancing Algorithm using two datacenters.

Fig. 7. Performance study of proposed Cat Swarm Optimization Load Balancing Algorithm with Particle Swarm Optimization Load Balancing Algorithm and Genetic Load Balancing Algorithm using three datacenters.

Fig. 8. Performance study of proposed Cat Swarm Optimization Load Balancing Algorithm with Particle Swarm Optimization Load Balancing Algorithm and Genetic Load Balancing Algorithm using four datacenters.
Genetic Load Balancing Algorithm using four data centers.

Table 6: Overall calculated Response Time of five datacenters

<table>
<thead>
<tr>
<th>SN</th>
<th>Configuration</th>
<th>Datacenter with VM</th>
<th>Response Time of Cat Swarm Optimization Load Balancing Algorithm</th>
<th>Response Time of Particle Swarm Optimization Load Balancing Algorithm</th>
<th>Response Time of Genetic Load Balancing Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C1</td>
<td>DC1,DC2,DC3,DC4,DC5 with 25 VMs each</td>
<td>325.63</td>
<td>332.87</td>
<td>338.56</td>
</tr>
<tr>
<td>2</td>
<td>C2</td>
<td>DC1,DC2,DC3,DC4,DC5 with 50 VMs each</td>
<td>316.02</td>
<td>322.84</td>
<td>329.76</td>
</tr>
<tr>
<td>3</td>
<td>C3</td>
<td>DC1,DC2,DC3,DC4,DC5 with 75 VMs each</td>
<td>311.93</td>
<td>319.45</td>
<td>325.88</td>
</tr>
<tr>
<td>4</td>
<td>C4</td>
<td>DC1,DC2,DC3,DC4,DC5 with 25,50,75 VMs each</td>
<td>309.97</td>
<td>316.65</td>
<td>324.03</td>
</tr>
</tbody>
</table>

Fig. 9: Performance study of proposed Cat Swarm Optimization Load Balancing Algorithm with Particle Swarm Optimization Load Balancing Algorithm

Fig. 10: Performance study of proposed Cat Swarm Optimization Load Balancing Algorithm with Particle Swarm Optimization Load Balancing Algorithm and Genetic Load Balancing Algorithm using six data centers.

In a distributed balancing environment, the algorithm implemented might change by their size or topology so, it would lead to higher Cost. The virtual machine Cost and data transfer Cost together binds to form grand Cost. The primarily aim of load balancing is to reduce the Response Time in a reasonable Cost. So, by using a load balancing technique which would effectively balance the whole system with low Cost is most efficient. The simulation results depicting the Cost Details for Genetic Load Balancing Algorithm, Particle Swarm Optimization Load Balancing Algorithm and Cat Swarm Optimization Load Balancing Algorithm are mentioned in Table 8 and the performance analysis of Cost Details is represented graphically in Fig. 8.
Table 8: Overall calculated Cost Details of proposed Cat Swarm Optimization Load Balancing Algorithm with Particle Swarm Optimization Load Balancing Algorithm and Genetic Load Balancing Algorithm.

<table>
<thead>
<tr>
<th>S/No</th>
<th>Algorithm</th>
<th>VM Cost</th>
<th>Data Transfer Cost</th>
<th>Grand Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Cat Swarm Optimization Load Balancing Algorithm</td>
<td>0.402</td>
<td>2.323</td>
<td>2.725</td>
</tr>
<tr>
<td>2</td>
<td>Particle Swarm Optimization Load Balancing Algorithm</td>
<td>0.500</td>
<td>2.532</td>
<td>3.032</td>
</tr>
<tr>
<td>3</td>
<td>Genetic Load Balancing Algorithm</td>
<td>0.502</td>
<td>2.549</td>
<td>3.051</td>
</tr>
</tbody>
</table>

Fig. 11. Performance study of the Cost Details of proposed Cat Swarm Optimization Load Balancing Algorithm with Particle Swarm Optimization Load Balancing Algorithm and Genetic Load Balancing Algorithm.

5. CONCLUSION

In this paper, Cat Swarm Optimization Load Balancing Algorithm technique has been approached as a new way of balancing load by reducing the most vital parameters like Response Time and Cost Details. Deeply analyzing the simulation results we conclude that Cat Swarm Optimization Load Balancing Algorithm is the most efficient technique than Particle Swarm Optimization Load Balancing Algorithm and Genetic Load Balancing Algorithm. These three algorithms belong to dynamic environment and Genetic Load Balancing Algorithm is mainly concerned to centralized balancing whereas Particle Swarm Optimization Load Balancing Algorithm and Cat Swarm Optimization Load Balancing Algorithm are distributed balancing algorithms. As Cost Details play the most overriding role in Cloud Computing Environment, so Cat Swarm Optimization Load Balancing Algorithm resembles as the best among the other two i.e Particle Swarm Optimization Load Balancing Algorithm and Genetic Load Balancing Algorithm. Future work can be concerned on improvising Cat Swarm Optimization Load Balancing Algorithm by adding more processes.

REFERENCES

Social Engineering Threats and Applicable Countermeasures

A.O. Adewole
Department of Computer Science
The Polytechnic, Ibadan
Ibadan, Nigeria.
Olu_gold@yahoo.com

A.E Durosinmi
Department of Computer Engineering
Moshood Abiola Polytechnic
Abeokuta, Ogun State. Nigeria.
cunlexie@hotmail.com

ABSTRACT

The Information and Communication Technology (ICT) security in a socio-technical world was explored and focus made in particular on the susceptibility to social engineering attacks. Social engineering is the most commonly used tactic across all levels of adversaries to gain unauthorized access into a network. While many organizations attempt to implement a policy and technical capabilities to mitigate against this threat, network intrusions through social engineering attacks are often still highly successful. A proven way to assess an organization’s risk to these threats is to test the effectiveness of existing technical and organizational protections, starting with the security awareness of personnel. Most social engineering takes place via email, text message and phone. However, tactics can include simply walking in the front door behind someone possessing a valid badge, or dropping portable USB drives in the parking lot and waiting for an unsuspecting employee to plug them into their work computer. Whatever form social engineering takes, businesses and organizations are largely unprepared for how to effectively counter these attempts across their workforces. Getting employees’ attention and commitment to vigilance can be difficult without proving how easy those employees can be exploited. This paper explores this social engineering attack; analyze counter measures against the attack and makes recommendations on how it can be mitigated.

Keywords - Social engineering, threats, security, intrusion and attacks.

1. INTRODUCTION

Social engineering (SE) has been largely misunderstood, leading to many differing opinions on what social engineering is and how it works. Intruders and hackers are on the lookout for ways to gain access to valuable resources such as computer systems or corporate or personal information that can be used by them maliciously or for personal gain. Sometimes they get their chance when there are genuine gaps in the security that they can breach. Often times, in fact more often than one can guess, they get through because of human behaviors such as trust – when people are too trusting of others, or ignorance – people who are ignorant about the consequences of being careless with information. Social Engineering uses human error or weakness to gain access to any system despite the layers of defensive security controls that have been implemented via software or hardware.

The ultimate security wall is the human being, and if that person is duped, the gates are wide open for the intruder to take control [1][4]. Social engineering represents a type of confidence scheme aimed at gathering information, committing fraud, or gaining computer system access. Social engineering, almost by definition, capitalizes on human psychology, such as cognitive limitations and biases, which attackers exploit to deceive the victim. This differs from other types of UIT incidents, such as cases in which an individual inadvertently discloses sensitive information without any interaction with an outside party (e.g., posting information on public databases or losing information by discarding it without destroying it). The adversary (or adversaries) masterminding the social engineering UIT incidents may have one or more malicious objectives that correspond to the intended impact to the organization, such as financial loss, disruption, or information compromise [6][4]
Overview Of Social Engineering
There are two main categories under which all social engineering attempts could be classified – computer or technology based deception, and human based deception.

a. The technology-based approach is to deceive the user into believing that he is interacting with the ‘real’ computer system and get him to provide confidential information. For example, the user gets a popup window, informing him that the computer application has had a problem, and the user will need to re-authenticate in order to proceed. Once the user provides his id and password on that pop up window, the harm is done. The hacker who has created the popup now has the user’s id and password and can access the network and the computer system [3].

b. The human approach is done through deception, by taking advantage of the victim’s ignorance, and the natural human inclination to be helpful and liked. For example, the attacker impersonates a person with authority. He places a call to the help desk, and pretends to be a senior Manager, and says that he has forgotten his password and needs to get it reset right away. The help desk person resets the password and gives the new password to the person waiting at the other end of the phone. At the very least, the individual can now access the Personnel systems as if he were the manager, and obtain the social Security numbers and other confidential/private information of several employees. He could of course do more damage to the network itself since he now has access to it [2][1].

Fig 2: Frequency of social engineering when compared to other security Attacks.

2. STEPS IN SOCIAL ENGINEERING ATTACKS

Fig.1: Social Engineering Progression
In the first phase, information gathering, an attacker uses various techniques to track down detailed information that can be used to gain the trust of an individual connected to the targeted organization. The attacker will then use this information to develop a relationship with the individual in phase 2 of the attack cycle. This can take one phone call or it can happen over a period of weeks or even months. After the relationship is established, the attacker will exploit the relationship (phase 3) to get the target to reveal information or perform an action that would not otherwise take place. Phase 3 either accomplishes the attacker's objective or opens the door to achieving the final objective in phase 4.

2.1 Guarding Against Social Engineering

Social engineering attacks are elusive and underhanded. However, they are not impossible to combat, organizations need to implement processes that undermine the effects of social engineering and, beyond that, establish a culture of security and accountability within the company (Defense, Awareness, & Company, n.d.). One way to test the current security culture of an organization is to do a simple self-quiz. Think about how the employees in an organization would react if an unfamiliar person who looked out of place sat down in a cubicle and started working on a computer. The following questions should be asked:

i. Would one of the employees become suspicious about this event?
ii. Would any employee choose to report it?
iii. Would any employee know how to report it and who to report it to?

If an organization do not feel confident that employees would be able to intervene in this potential security breach, there is a need to take several concrete actions to improve organization’s security culture. Assuming a well-conceived security policy is in place, the first and most important action is to educate users about your company’s security policy, or at least the parts of it that potentially affect them.

3. HOW SOCIAL ENGINEERING WORK

Social engineering is defined as a “non-technical kind of intrusion that relies heavily on human interaction and often involves tricking other people to break normal security procedures.”

3.1 Common Social Engineering Scenarios Include:

Employee Information: Documents that contain information about employee’s names, departments etc. are very important as they can be used during the physical penetration test as information which is valid. Knowing already information from inside will allow you to establish more easily the trust as you will appear as someone valid.

- Emails: Obviously you can find corporate emails and from other sources like LinkedIn, official website etc. but also papers containing some email address is always a good finding as you will be able to discover internal information and also the structure of the emails accounts inside the company.

- Headed Papers: these kinds of papers can help penetration testers to create forgeries of the documents. This is essential for any social engineering engagement as you can cheat the employees to perform the action that you want.
What follows are specific measures through which users and organizations can mitigate against social engineering attacks.

1. **Skepticism is Healthy**: No information without verification! Do not provide any personal or confidential information over phone, text, or internet to anyone unless you can verify who that person is and that person actually has a legitimate need for the said information. Employees are often scammed into revealing sensitive information by social engineers who pretend to be IT professionals from the same company. Dispose of any sensitive documents with shredders, keep your computer protected with anti-virus programs, and most importantly of all, don’t be gullible and thus get tricked into sharing confidential information. Remember that skepticism is a good thing.

2. **Check your Status**: There are plenty of security agencies that companies and individual contract just to protect them against the threat of social engineering. These agencies can gauge how vulnerable your network or organization is to social engineering attack. This can often be a wake-up call for many companies as well as individuals.

3. **No ‘Phishy’ business**: ‘Phishing’ is a very popular method of social engineering. E-mails requesting personal information is sent to people from seemingly legitimate sources (banks, financial organizations etc.) To inspire confidence and create a sense of false security. Sometimes these e-mails redirect people to fake websites that closely resemble the original and then proceed to extract personal data. ‘Pharming’ is another such method that redirects people to fake websites nearly identical to the legitimate one they are trying to access. There are several security software programs that combat Phishing and pharming. But make sure your network’s employees are security conscious and aware of such scams because there is no substitute for being plain vigilant.

4. **Use the right software**: Firewalls and anti-virus programs are very important for any network to use for obvious reasons. But these days content filtering systems and programs are becoming increasingly popular. They increase online security by blocking malicious websites and prevent users to becoming prey to phishing and pharming. In addition to this, you should never forget to keep your system software up to date. Patches and updates often fix security loopholes.

5. **Security Awareness**: A culture of security awareness can go a long way and it is of the utmost importance in any organization or company or network. Most people do not fall prey to such attacks intentionally. Both executives and employees should be educated on basic security training to enable them to protect confidential data. In fact, executives are more vulnerable because they have a relative lax attitude towards security protocols. Implement basic security measures to protect confidential data like classification of sensitive information and two-factor authentication for sensitive data. This can help make your network nearly impermeable.

5. **CONCLUDING REMARKS**

In this paper, we x-rayed social Engineering in its many guises. Using concrete examples, we showed how social engineering scams can be used to defraud unsuspecting users. We also revealed that social engineering scams can occur via email, websites, text messages, and sometimes phone calls. We embellished the social engineering scenario in order to provide some understanding to electronic mailing systems and online consumers that can serve as bases for empowering users and organization in their quest to mitigate social engineering attacks.

**REFERENCES**


Intelligent Allocation of E-Learning Resources Using Tokenization and Crawling Algorithms

Y. Folajimi & E.E. Edet
Department of Computer Science
University of Ibadan
Ibadan, Nigeria
yetundeofolajimi@gmail.com; edet_e_emmanuel@yahoo.com

ABSTRACT

This paper examines the challenges of gaining access to relevant learning materials in learning management system (LMS) platforms by learners; which includes lack of tool to access digital materials that are relevant to the course of study or learning goals, periodic availability, lack of high scalability to meet the exponential growth of users, educational materials and services. We propose a framework that meeting their students’ learning goals to an acceptable degree. Our framework, Intelligent e-Learning Resource Allocation (IneLRA), adopted tokenization method and crawling algorithm. The user’s search term is sieved into tokens and based on the user’s profile, preference and device; the system crawls the cloud and a customized result based on page ranking algorithms is provided to the user. The method includes a web-based embedded model of content adaptation engine, a slate interface, a data miner and a controller in the e-learning system. Evaluation results indicate significant improvement on the learning efficiency of users in meeting their learning goal(s) measured on the metrics of search time, search volume and relevance of search result.

Keywords: tokenization, crawling algorithm, online learning, Artificial intelligence, e-Learning, Cloud Computing.

1. INTRODUCTION

With the huge growth of the number of students, education contents, services that can be offered and resources made available, e-Learning system dimensions grow at an exponential rate. The challenges of optimizing resource computation, storage and communication requirements, and dealing with dynamic concurrency requests highlight the necessity of the use of a platform that meets scalable demands and cost control. This environment is Cloud Computing. The integration of cloud computing to an E-Learning system is what is known as Cloud Based E-Learning (CBeL) or education Software as a Service (SaaS); a web based learning system that support activities used in achieving learning goal(s). In clear connection with this integrated platform system, several educational institutions have implemented these systems already [1] and these are seen in the numerous number of Learning Management Systems (LMSs) that are in use. For instance, 27% of schools make use of cloud facilities in the USA [2]. Other programmes such as that in the UK are designed to provide a private education cloud with appropriate data management, storage and tools [3].

These LMSs come with various features, particularly test scoring, testing, training metrics, training tracks, user access controls, users size served, virtual classes, waiting listing, performance assessment, exam engine, registration management, etc. but no tool that allows students to search for materials on subject matter that will aid their meeting learning goal(s).

In this work, we propose an intuitive approach called Intelligent e-Learning Resource Allocation (IneLRA) model to allocate relevant digital artefacts that will meet learning goal(s) of students in a cloud based e-learning platform.

2. RELATED WORKS

Cloud computing is a new paradigm of computing that is being employed by several organizations and institutions today. The conception is derived from the description of the “Internet cloud” whereby the imagery of a cloud is conventionally “used to represent the Internet or a large networked environment” [4]. The concept owes much to the evolution of infrastructures based upon client-server, application service provider (ASP), service oriented architecture (SOA), grid computing and even “time -slicing” of mainframe computers, virtualization, autonomic system computing, the internet, Web services, web application frameworks, parallel computing and open source software. [5].
Service Oriented architecture (SOA) facilitates the availability of cloud computing services as web services which is accessed through the Internet. SOA also makes it viable for cloud computing services to be accessible on multiple platforms [6]. Given that cloud computing services are web based application that gains access via the Internet, Web 2.0 provides cloud computing with capabilities of enhanced connectivity and communication between web applications. This makes the access to cloud computing services by users more proficient and less complex [6]. Grid computing enables cloud computing with resource sharing and control capability.

Cloud computing presages a major transformation of how we store information and run applications. Rather than running programs and storing data on stand-alone systems, these programs and data are all hosted in the “cloud” - an indistinct collection of computers and servers accessed via the internet, or the collection of several data-centers of hardware and software.[7].

There is presently no common standard or definition for cloud computing and several authors have come up with so many definitions that it makes it difficult to ascertain which definition best describes the cloud computing concept. A comprehensive survey conducted in 2009 by the University of California Berkeley RAD Lab (Reliable Adaptive Distributed Systems Laboratory) conceded a definition that has been gaining across-the-board recognition: “Cloud Computing refers to both the applications delivered as services over the Internet and the hardware and systems software in the datacenters that provide those services. The services themselves have long been named as Software as a Service (SaaS).

In a more comprehensive perspective, the National Institute of Standards and Technology (NIST) defines cloud computing as “a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., Networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction”[8]. In a nutshell, the idea is that this is an accessible resource of hardware and software which an organization or individual can harness, anywhere in the world via the internet.

The aim of Cloud Computing is to enable enterprise search with the capability to further prune costs through enhanced implementation, trim maintenance and infrastructure cost and enhance deployment cycles.

### 2.1 Integration Model

A general framework which allows for a detailed characteristics for managing a virtual computing laboratory was proposed in [9], but this architecture is not tailored to adjust to the needs of an e-learning system. Adaptivity in the e-learning system is realized using the model described in [10]. Group personalization of courses has been implemented with respect to Felder Silverman learning styles. Different learning materials and activities are created for each group. All student activities are tracked using the Moodle LMS. Personalization is reflected in adapted resources, activities, tasks and tools for communication.

Integrating e-learning services with the cloud computing infrastructure includes: the implementation of LMS, the integration of Internet services in a network of educational institutions and a business information system, with the use of multilayer: Human resource integration – (students, teachers interacting together); Information integration – (gathering and converting heterogeneous and unstructured data to structured data); Process integration and Application integration and the methods for integration included: envisioning, planning, developing, stabilizing and deploying phases. [10]

### 3. PROPOSED MODEL

In order to improve the learning efficiency and consequently providing a method that is efficient and reliable enough for user to achieve their learning goal(s), the e-learning platform has continuously been improved upon by several researchers; and this has led to the development of more functional and stable learning environment.

Learners’ ability to access digital materials (e-books, lecture videos, PowerPoint slide shows etc.) that are relevant to the learner’s learning course will further encourage the learner in the learning process, forms the aim of our methodology. The IneLRA model as shown in Figure 1 consists of modules that achieve effective learning process in order to achieve learning goal(s). These include:

- Content Adaptation Engine (CAE)
- The Moodle (LMS)
- Cloud Infrastructure
- Data Miner (DM)
- Slate Interface (SI)
- Web User Interface (WUI)
- Controller
Functions of System Components
The User interacts with the intelligent resource allocation e-Learning system through the Web User Interface (WUI) on the user’s device; after authenticating the user, the CONTROLLER module sends a request to the CONTENT ADAPTATION ENGINE (CAE) to ascertain the type of device and the feedback is sent to the controller. The controller, based on the feedback prompts the DATA-MINER (DM) for educational materials.

The Data-miner sources for materials from the cloud, based on the information it has been fed with by the controller. The search result is then held-up in a container known as the SLATE INTERFACE (SI), which now alerts the controller of the completion of the search. The controller flags the search result on the screen (WUI) for the user. Consequent upon this, the following explains the individual function(s) of the components as well as the sub-components that make-up the system.

Web User Interface (WUI)
The Web User Interface (WUI) is responsible for user interaction with the intelligent resource allocation e-Learning system. It is the interface where a learner enters his/her credentials to access the platform. It underpins the activities of other interfaces in the system. The WUI provides a user-friendly interface to the user and act as a communication medium between the teacher, the learner and the system. The user accesses the system through a web-browser.

Controller
The Controller is like the Central Processing Module of the system. It prompts for information and forwards same to the appropriate component for processing and the feedback is again returned to it for display or further processing.

The controller waits for a search request from the user, and then captures the user’s search keywords, and couple with the information about the user’s device, preference and immediate previous activities retrieved from the CAE, the controller alerts the data-miner and wait for feedback before presenting the search result to the user.

Content Adaptation Engine (CAE)
The Content Adaptation Engine (CAE) is a compound module; it consists of two (2) sub-components, namely:
- Device Detection Engine (DDE) embeds WURFL framework,
- Context Detection Engine (CDE)

Device Detection Engine (DDE)
Device detection can be achieved when the access device makes a HTTP request. The device profile is stored on the server of the manufacturers. The DDE embeds the WURFL (Wireless Universal Resource File) device detection framework is used because of its regularly updated feature [11]. The diagrammatical description of the detection process flow for a device (desktop or mobile) proposed by [2] is shown in figure 2.

![Figure 1: Intelligent e-Learning Resource Allocation (IneLRA) Architecture](image)
When a match of the user device is found from the WURFL device repository, by merging the information from the device discovery and that of WURFL repository, the information, which contain both the software and hardware feature information of the user’s device is then sent alongside the user’s contextual information for further usage.

Hardware Information that is retrieved includes:
- Device name,
- CPU processing power,
- memory size,
- screen resolution
- Screen size in pixels
- The existence of the touch screen
- Support for certain graphics formats

Software information that can be retrieved includes:
- the device operating system,
- Preferred markup language
- browser capability
- Support for certain multimedia files and Java
- Support for styles

Context Detection Engine (CDE)
The Context Detection Engine (CDE) collects contextual data about the learner, estimates the context, classify, filter and prioritize the data. The following are the contextual information of the user:
- Network bandwidth
- Learning activities (subject, recent activities)
- Neighbours (other learners and Current Learning activities)
- User preference (learner’s goal and Interest, settings on mode of receiving information- graphical or text). The user preference is determined by monitoring the behavior of the Learner from the interaction with the system, and pop-up questions that requires the user to choose from options.

The resulting information from both the DDE and the CDE are passed on from the CAE to the controller. And again, the controller prompts the DM to fetch the information, which will then be held in the Slate interface. Any information with features that are within the capability of the device as a whole will be displayed; otherwise, it will not be displayed.

Data-Miner (DM)
The Data-Miner (DM) receives a search keyword from the controller and couple with the other information received from the CAE about the user’s device and preference, a search is carried out to retrieve materials that will meet the user’s learning process.
The crawling algorithm is as shown in figure 3.

```c
void crawl()
{
    while (NextURLIDScanned < NextURLID) {
        urlIndex = NextURLIDScanned
        Fetch the url1 entry in urlIndex
        NextURLIDScanned++
        Get the first 100 characters or less of the document from url1 without tags. Add this
description to the URL record in the URL table.
        For each url2 in the links in the anchor tags of this document {
            fetch the url2 in the link
            if it is not text/html continue;
            if (NextURLID < MaxURLs && url2 is not already in URL table) {
                put (NextURLID, url2) in the URL table
                NextURLID++
            }
            Get the document in url1 without tags
            for each different word in the document {
                In Word Table create a new (word, URLID) if the entry does not exist.     }
        }
    }
}
```

Figure 3: Web Crawler Algorithm

The crawler begins with the seed sites and scans the links in
the site up to a certain depth as specified by the programmer,
in our case, we used depth-3. The visited links are marked
and the crawler moves on to the next sites on the list. A
record of the sites visited is taken and the collated links are
ranked based on the weight of the link in terms of the amount
of hits it has, amongst other measuring factors.

**Slate Interface (SI)**
The slate Interface is simply a temporary memory location
that holds the list of search result as compiled from the pool
of heterogeneous sources on the Internet.

**Cloud Infrastructure**
The main components of the cloud infrastructure include:
services for access to a virtual environment, services for
resource management, services for user account management,
a system for distributed file management, and a virtual
infrastructure management system. These components allow
for an effective work with a VM. VMs are stored in the
image repository and can be moved and run on users’
demands [13].

The web services form the platform through which the LMS,
OpenNebula and OpenLDAP interact. The user accounts are
stored on LDAP server. The LDAP server is integrate
d with the user directory of the educational institution where the
student accounts are located. User authentication is
performed using the LDAP protocol.

The same login information is used to log on to the LMS.
OpenNebula allows users to attach/detach volumes to their
instances.

### 4. IMPLEMENTATION AND RESULT DISCUSSION

Usually users sourcing for materials use search engines such
as google, yahoo, AOL, mamma, ask etc, and the
performance of which has been reasonable. The
implementation which followed a bottom-up approach used
tokenization technique and crawling algorithm to search and
page ranking algorithm to sort the search result in order of
relevance. The result which was measured with conventional
search result was used to determine the improvement on the
learning efficiency of users. First, we implemented a search
engine that crawls the internet (with some major sites as seed
sites, namely academia, youtube, digitalbooks, yahoo,
wikipedia) without preference to the user’s profile
in use. We then implemented an upgrade search engine that
filters search result based on the user’s profile and
preference; with the expectation that the search result in
terms of volume will be lesser.

The user’s device was an additional result extraction factor
that was included in order to deliver customized digital
materials to mobile devices. In order to handle the aspect of
scalability and round-the-clock accessibility, we hosted the
site on the Amazon EC2.
5. RESULT DISCUSSION AND EVALUATION

The use of the system for different search terms showed a reduction in search time and volume of search result and an increase in the search result relevance when the same search term is queried in other search engines (Google, Yahoo, bing, aol).

Three estimation metrics were used in evaluating the performance of the system against the conventional search engines. They are:

- Search Time (ST)
- Search Volume (SV)
- Search Relevance (SR)

**Search Time comparison**

An important factor in evaluating the system’s performance is the search time. We compared the search time in seconds of ten search instances against the conventional search engines. We used the same search term at each search instance across the search engines. This is to provide a fair basis for our conclusions and claims. We claim that our approach have saved an approximate of 38% of search run time. It is a general and obvious consensus that the search time is uniformly reduced across search terms and devices type, as shown in figure 5 without consideration to the network quality.
Our approach, and thus, the implementation are not just better in search time reduction, it is better in storage space reduction also. Generally the search result gotten from each instance of search, when compared to the conventional search engines were lesser as shown in figure 6. This is because of the filtering feature based on user’s preference and profile as well as user’s device.
Search Relevance Comparison

In order to meet the aim of this research, it was needful to estimate the degree of relevance of the search result in meeting learning goal(s). A Page ranking (PageRank) algorithm was adopted to determine a page relevance to a search query by assigning numerical weightings to hyperlinked documents (or web pages) indexed by the system. The Page Rank value of a page reflects the frequency of hits on that page by the random surfer. A hyperlink to a page counts as a vote of support.

\[
\text{PageRank}(p_i) = \frac{q}{N} + (1 - q) \sum_{p_j \in L(p_i)} \text{PageRank}(p_j)
\]

where \( p_1, p_2, \ldots, p_N \) are the pages under consideration, \( L(p_i) \) is the set of pages that link to \( p_i \), and \( N \) is the total number of pages. For the purpose of this research, we computed the Relevance Percentage (RP) of search result by dividing the number of the first page with the total volume of search result and multiplied by 100. Ten (10) was taken as the benchmarks for the number of search items for each first page as adopted by most popular search engines such as google, ask, msn etc, which refers to the ten most relevant search results based on the page ranking by the search engine.

A graphical representation of the search volume comparison between IneLRA model and other conventional search engines is shown in figure 7.

![Search Relevance Comparison](image)

Figure 7: Search Relevance comparison

It was observed from the three metrics used to evaluate this approach that the Search Time (ST) is inversely proportional to the Search Volume (SV) and the search result relevance decreases as the search volume increases.

\[
ST \propto \frac{1}{SV}
\]

\[
SR \propto \frac{1}{SV}
\]

The Search time reduced by 38.09%, the Search volume reduced by 76.07% and the Search relevance improved by 61%. The graphical representation is shown in figure 8.
6. SUMMARY AND CONCLUSION

This work has proposed an Intelligent e-Learning Resource Allocation (IneLRA) approach to accessing digital materials on a Learning Management System (LMS) platform by using tokenization and crawling technique and filtering the pages based on user’s preference, profile and device type; the result of which was then ranked using a PageRank algorithm. Results of the implementation indicate significant improvement on the learning efficiency of users in meeting their learning goal(s) measured on the metrics of search time, search volume and relevance of search result. It can be concluded that our approach improves learning efficiency significantly when compared with conventional search engines. Notably, the approach gives better search runtime (38%) and far lesser search volume (76%) with a significant degree of relevance (61%) that can be perused quickly by users in meeting their learning goals.
REFERENCES

Harmonic Analysis in a 33kV Distribution Network: A Case Study of Island Business District

L.M. Adesina
Department of Inspection and Quality Assurance
Eko Electricity Distribution Plc (EKEDP)
Lagos – Nigeria
lambe.adesina@ekedc.com; lmamaj@yahoo.com; Mobile: 08023091620

O.A. Fakolujo
Department of Electrical/Electronic Engineering
University of Ibadan
Ibadan – Nigeria
ao.fakolujo@ui.edu.ng; olaosebikanfakolujo@gmail.com

ABSTRACT

Distribution Line Network harmonic investigation has become an important issue in electric power systems since the increased use of power electronic devices and equipment sensitive to harmonics, which in turn, has increased the number of adverse harmonic-related events. Power quality problem and the means of keeping it under control is a growing concern. Due to the connection of nonlinear loads in a 33kV distribution network, power quality issues have been raised and have resulted to widespread waveform distortion. This paper makes use of the Busbar voltage solutions obtained from a power flow study on the Eko Electricity Distribution Company’s 33kV network and they are used as the base data for the harmonic analysis. A distribution line model was formulated. MATLAB/Simulink was used to simulate this modelled network for each distribution lines of the power system network considered. This simulation was achieved through the application of sending and receiving end busbar voltages already obtained from the power flow study. The harmonics observed on each distribution line at various scenarios of load/feeder restoration and their characteristic impedance magnitudes/frequency are all presented. Necessary discussions on the harmonic results are also presented.

Keywords: Distribution Line, Power Quality, Harmonics

African Journal of Computing & ICT Reference Format:

1. INTRODUCTION

The late 80’s and early 90’s saw the emergence of power electronic loads. Power electronic loads offer the advantages of efficiency and controllability. However, they draw non-sinusoidal currents from AC power systems, and these currents react with system impedances to create voltage harmonics and in some cases, resonance. Studies have shown that harmonic distortion levels in distribution systems are rising as power electronic devices continue to proliferate [1]. Distribution Line network harmonic investigation has become an important issue in electric power systems since the increased use of these devices and equipment sensitive to harmonics, which has increased the number of undesirable harmonic-related conditions. The Power quality problem, and the means of keeping it under control, is a growing concern [2]. Power quality means the physical characteristics of the electrical supply provided under normal operating conditions that do not disturb the customer’s processes [3].

Power Quality has no fully accepted definition, but surely the response involves the waveforms of current and voltage in an ac system, the presence of harmonic signals in bus voltages and load currents, the presence of spikes and momentary low voltages, and other issues of distortion [4]. Perhaps the best definition of power quality is [5], “the provision of voltages and system design so that the user of electric power can utilize electric energy from the distribution system successfully, without interference or interruption”. The non-linear characteristics of power electronic loads cause harmonic currents, which result in additional losses in distribution system equipment, interference with communication systems, and misoperation of control [2].
Harmonics are components of a periodic wave having a frequency that is an integral multiple of the fundamental power line frequency. Harmonics, therefore, are the multiple of the fundamental frequency. Traditional harmonic producing loads are the rotating machines and transformers in electric power systems. Because of uneven distribution of flux in the air gap of the rotating machine, non-sinusoidal voltages and currents are generated by rotating machines such as synchronous machines. Overloaded transformers generate harmonic currents. Transformer excitation currents will be non-sinusoidal because of core-saturation, magnetic imbalance or DC magnetization resulting in harmonic currents [6, 7]. The increasing usage of non-linear loads on electrical power systems is causing greater concern for the possible loss of transformer life[8,9].

Small power equipments such as televisions, computers, Adjustable Speed Drives (ASD), Uninterrupted Power Supply (UPS) systems, electronic ballasts and office equipments such as printers and fax machines draw non-sinusoidal voltages and currents. The current harmonics consumed by these loads flowing through the line impedances cause distorted voltages and, the significant harmonics are extended to the rest of the network [10]. Individual effects of this type of equipment to harmonic distortion may be negligible, but the combined effects of a larger number of these to harmonic distortion in power systems can be substantial. The network was modelled and simulated using the MATLAB/Simulink platform. Harmonic analysis in general can benefit from effective monitoring [11] of a power system network.

2. SOURCES OF HARMONICS

In order to understand the injection of harmonic currents in the power distribution network, it is necessary to discuss the general characteristics of non-linear loads. Nonlinear loads inject harmonic currents or voltages into the distribution network even when fed by a sinusoidal voltage or current waveform. Nonlinear loads can be broadly divided into two categories [12]:

- Harmonic source type loads;
- Harmonic voltage source type loads.

The associated harmonic current passing through the system impedance causes voltage drops for each harmonic frequency based on Ohm’s Law. The vector sum of all the individual voltage drops results in total voltage distortion, the magnitude of which depends on the system impedance, available system impedance, available system fault current levels of harmonic currents at each harmonic frequency [13].

The effects of harmonics can be summarized as follows:

- High fault current (stiff system)
- Distribution system impedance and distortion is low.
- Harmonic current draw is high.
- Low fault current (soft system)
- Distribution system impedance and distortion is high
- Harmonic current draw is low

For general purposes, the harmonic sources can be divided into three categories [14]:

1. A large number of distributed nonlinear components of small rating (i.e. mass products), consists mainly of: single phase diode bridge rectifiers, power supplies of low voltage appliances (SMPS in TV sets, PCs and other IT equipment), and gas discharge lamps.
2. Large static power converters (SPC) are used more extensively for controlling loads. There are many forms of SPC: rectifiers, inverters, cyclo-converters, single-phase, three-phase, twelve-pulse, six-pulse, but all have the same character. They are all nonlinear and they inject non-sinusoidal current into the power system.
3. Large and continuously varying nonlinear loads. This refers mainly to electric metal-melting arc furnaces with power ratings in the tens of megawatts and connected directly to the transmission network. The furnace arc impedance varies randomly and extremely asymmetrical, since the carbon electrodes in contact with iron have dissimilar impedances between the positive and negative flows of current. Resistance welding has the same characteristics, where the copper electrodes and the steel being welded have dissimilar impedances between the positive and negative flows of current.

2.1 Impact of Harmonics

Practically speaking, utilities frequently encounter the following harmonic-related problems [12]:

- Voltage distortion in distribution feeders;
- Increased RMS currents, heating and line losses;
- Overheating of power transformers, which requires higher K-factor transformers;
- Derating of distribution equipment;
- Overloading of phase and neutral conductors - neutral currents in a typical commercial office building may carry more than phase RMS currents;
- Overloading and fuse blowing of power factor correction capacitors;
- Tripping of voltage harmonic sensitive equipment;
- Failure of control electronics, micro-processors;
- Reduced accuracy of measuring instruments (such as watt-hour meters);
- Malfunction of solid-state fuses, breakers and relays;
- Reactive power and resonance problems;
- Reduced system stability and safe operating margins, etc.
2.2 Harmonic Components

A method to represent any non-sinusoidal periodic function \( u(t) \) using an infinite series of cosine and sine functions and coefficients as shown in equation (1) was first proposed by Baron Jean Fourier in 1822:

\[
    u(t) = A_0 + \sum_{n=1}^{\infty} \left[ A_n \cos(n\omega_0 t) + B_n \sin(n\omega_0 t) \right]
\]

\[
    = A_0 + \sum_{n=1}^{\infty} C_n \cos(n\omega_0 t + \Psi_n) \quad \text{..................(1)}
\]

Where

\( u(t) \) is a periodic function of frequency \( f_0 \), angular frequency \( \omega_0 = 2\pi f_0 \), and period \( T = \frac{1}{f_0} = \frac{2\pi}{\omega_0} \).

\( C_n \cos(n\omega_0 t + \Psi_n) \) represents the fundamental component, and \( C_n \cos(n\omega_0 t + \Psi_n) \) represents the \( n \)th harmonic component of amplitude \( C_n \), frequency \( n\omega_0 \), and phase \( \Psi_n \) relative to the fundamental.

Generally, for power systems, the fundamental frequency is either 50Hz or 60Hz. Power systems in Nigeria are typically operated at 50Hz and thus harmonic frequencies will appear as multiples of 50Hz (100Hz, 150Hz, 200Hz, etc.). The Fourier series coefficients \( C_1, C_2, \ldots, C_n \) and relative phases \( \Psi_1, \Psi_2, \ldots, \Psi_n \) make up the harmonic spectrum of the waveform and are found using equations (2) through (6):

\[
    A_0 = \frac{1}{T} \int_0^T u(t) \, dt = \frac{1}{\omega_0} \int_{2\pi}^{2\pi} u(\theta) \, d\theta \quad \text{where} \quad \omega_0 = \frac{2\pi}{T} \quad \text{............(2)}
\]

\[
    A_n = \frac{1}{T} \int_0^T u(t) \cos(n\omega_0 t) \, dt = \frac{1}{\omega_0} \int_{2\pi}^{2\pi} u(\theta) \cos(n\theta) \, d\theta \quad \text{............(3)}
\]

\[
    B_n = \frac{1}{T} \int_0^T u(t) \sin(n\omega_0 t) \, dt = \frac{1}{\omega_0} \int_{2\pi}^{2\pi} u(\theta) \sin(n\theta) \, d\theta \quad \text{............(4)}
\]

\[
    C_n = \sqrt{A_n^2 + B_n^2} \quad \text{..........................(5)}
\]

\[
    \Psi_n = \tan^{-1} \frac{B_n}{A_n} \quad \text{..........................(6)}
\]

Conversely, if the harmonic spectrum of a given current or voltage waveform \( u(t) \) is known the original waveform can be constructed using the Fourier series summation:

\[
    u(t) = \sum_{n=1}^{\infty} U_n \cos(n\omega_0 t + \Psi_n) \quad \text{..........................(7)}
\]

where

\( U_n \) is the \( n \)th harmonic peak current or voltage, \( \Psi_n \) is the \( n \)th harmonic phase, \( \omega_0 \) is the fundamental angular frequency,
\( \omega_0 = 2\pi f_0 \) and \( f_0 \) is the fundamental frequency, typically 50Hz.

3. HARMONIC SIMULATION

In order to study the net effects of harmonics in a power system, simulations were carried out taking the 33kV distribution network of Island Business Unit (IBU), Eko Electricity Distribution Plc (EKEDP) grid, as a case study and the network is shown in fig. 1.

In power systems, transmission or distribution lines are modelled using any of these networks: \( \pi \) - network, \( T \) – network and \( \pi \) & \( T \) networks, depending on the distance of the line.

Since the longest distribution route length in this journal paper is 6.84KM, a \( \pi \) distribution line modelling approach was adopted using the MATLAB/Simulink environment. The \( \pi \)-modelled diagram or circuit suitable for the display of harmonics flowing between two buses is shown in fig. 2 below. It comprises the source, which is the input, as well as the output electrical parameters such as resistance, reactance, and susceptance. It also has two oscilloscopes placed at Fowler (FOW) and Alagbon (ALG) busbars (within the distribution network under consideration) to monitor the input and output voltage variation during simulation process. A reactor is connected for power control, while a current source is included to act as non-linear loads which generate harmonics.
3.1 Development of Flowchart for Harmonic Evaluation on the Line between Two Buses

The Busbar voltage solution obtained from power flow study of fig.1 is further applied on the modelled π distribution network to get the harmonics on the line. It involves the use of two voltages at the ends of the distribution lines i.e. sending end voltage, $V_S$ and the receiving voltages $V_R$.

The flow chart in figure 2 shows the procedure involved in obtaining harmonic flow on the line between two buses.
3.1 Harmonics Simulation Algorithms
The procedure involved in determining the Harmonic flow in the power systems Network using MATLAB simulation approach (Cascaded pi Distribution Model) are as follows:
1. Obtain a suitable Model Network as shown in Fig.1
2. For the line in consideration, insert system frequency and the sending end voltage obtained from power flow analysis results into the MATLAB block parameters of AC voltage source.
3. Display the Sending end Voltage (SV) signal on scope b (Input).
4. On MATLAB Shunt reactor (110Mvar) block parameters, input- Nominal Voltage $V_n$ (rms)
5. Nominal Frequency (50Hz) Active Power $P(w)$
6. Inductive reactive power $Q_L$ (positive var)
7. Vary the parameter in 4 above (i.e. Nominal Voltage $V_n$) until an exact corresponding Receiving end Voltage (RV) is achieved via scope o.
8. When step 5 is achieved, display the bus bar receiving end voltage signal (output) on scope o.
9. In the state-space model block, compute distribution line steady state parameters in magnitude and phase values. The parameters to be displayed will include values of:

   a) States:
      i. Current I through $Z_i$
      ii. Voltage $V_{d}$ across $Z_i$ (RLC)
      iii. Current I through $Z_o$
      iv. Voltage $V_{o}$ across $Z_o$
      v. Current I through shunt Reactor 110MVar
     vi. Input Voltage $V_{c}$ at the beginning of Distribution line (Bus bar 1)
     vii. Current I through the pi section (Distribution line)
     viii. Output voltage $V_{c}$ at the end of Distribution line (Bus bar 2)

   b) Measurements:
      Magnitude and phase values of:
      i. Measured voltage $V_o$ at scope o (output)
      ii. Measured voltage $V_b$ at scope b (Input)

   c) Sources:
      Magnitude and phase values of:
      i. Bus bar voltage in volts
      ii. Load on Bus bar in Amperes

1. Plot the power system impedance against frequency to display the harmonics on the distribution line.
2. From the plot, record harmonic frequencies and the corresponding harmonic impedance.
3. Repeat step 2-9 for other lines within the same time scenario. In these cases, the parameters (SE Voltage, RE Voltage and distance) may be different in values.
4. Prepare a table of Distribution line Harmonics and their impedance.
5. Go to “Next hour scenario”
6. Repeat steps 2-11
7. End
3.2 Harmonic Simulation Results

Figures 3a to figure 3m are the harmonic analysis results for different cases and at different times of the day under consideration. The amplitude of Source voltage, V (33kV), Nominal frequency (50Hz), Shunt Reactor (110MVar), Active Power, P (110e6/300 W) and Inductive Reactive Power, Q
\[ Q_L \] (110e6 +Var), remain the same in all cases of the simulation, except for the case of Alagbon – Anifowoshe D/L at 23:00Hrs, whose amplitude of source voltage is 32.825kV. Tables 1 to 4 give summaries of the Distribution Line Harmonics and their Impedance for all cases under consideration.

At 02:00Hrs
Case 1: Alagbon (ALG) – Alagbon Local (ALG/L)
Distribution Line (D/L)
Nominal Voltage, V (Vrms) = 39.5kV

Case 2: ALG – Fowler (FOW) D/L
Nominal Voltage, V (Vrms) = 91.13kV

Table 1: Distribution Line Harmonics and their Impedance for cases of 02:00Hrs

<table>
<thead>
<tr>
<th>D/L</th>
<th>Lnth (km)</th>
<th>SV (kV)</th>
<th>RV (kV)</th>
<th>Harmonics</th>
<th>Harmonic Imp. (Ω)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALG-ADM</td>
<td>6.13</td>
<td>33</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ALG-ALG/L</td>
<td>0.15</td>
<td>33</td>
<td>32.997</td>
<td>2\textsuperscript{nd}</td>
<td>10\textsuperscript{1.881}</td>
</tr>
<tr>
<td>ALG-ANI</td>
<td>6.84</td>
<td>33</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ALG-FOW</td>
<td>3.00</td>
<td>33</td>
<td>32.881</td>
<td>2\textsuperscript{nd}, 6\textsuperscript{th}, 15\textsuperscript{th}, 24\textsuperscript{th}</td>
<td>10\textsuperscript{2.0}, 10\textsuperscript{2.51}, 10\textsuperscript{2.2}</td>
</tr>
<tr>
<td>ADM-MAR</td>
<td>1.80</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ADM-ANI</td>
<td>2.40</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ALG-BAN/I</td>
<td>5.00</td>
<td>33</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

At 09:00Hrs
Case 1: ALG – Ademola (ADM) D/L
Nominal Voltage, V (Vrms) = 117.5kV

Case 2: ALG –ALG/L D/L
Nominal Voltage, V (Vrms) = 39.6kV

Table 1: Distribution Line Harmonics and their Impedance for cases of 09:00Hrs

<table>
<thead>
<tr>
<th>D/L</th>
<th>Lnth (km)</th>
<th>SV (kV)</th>
<th>RV (kV)</th>
<th>Harmonics</th>
<th>Harmonic Imp. (Ω)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALG-ADM</td>
<td>6.13</td>
<td>33</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ALG-ALG/L</td>
<td>0.15</td>
<td>33</td>
<td>32.997</td>
<td>2\textsuperscript{nd}</td>
<td>10\textsuperscript{1.881}</td>
</tr>
<tr>
<td>ALG-ANI</td>
<td>6.84</td>
<td>33</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ALG-FOW</td>
<td>3.00</td>
<td>33</td>
<td>32.881</td>
<td>2\textsuperscript{nd}, 6\textsuperscript{th}, 15\textsuperscript{th}, 24\textsuperscript{th}</td>
<td>10\textsuperscript{2.0}, 10\textsuperscript{2.51}, 10\textsuperscript{2.2}</td>
</tr>
<tr>
<td>ADM-MAR</td>
<td>1.80</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ADM-ANI</td>
<td>2.40</td>
<td>0</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ALG-BAN/I</td>
<td>5.00</td>
<td>33</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Case 3: ALG – Anifowoshe (ANI) D/L
Nominal Voltage, V (Vrms) = 116.5kV

Figure 3e: Harmonics on ALG – ANI D/L

Case 4: Alagbon – Fowler D/L
Nominal Voltage, V (Vrms) = 91.1kV

Figure 3f: Harmonics on ALG – FOW D/L

Case 5: ADM – ANI D/L
Nominal Voltage, V (Vrms) = 84kV

Figure 3g: Harmonics on ADM – ANI D/L

Case 6: ALG – Banana Island (BAN/I) D/L
Nominal Voltage, V (Vrms) = 114.4kV

Figure 3h: Harmonics on ALG – BAN/I D/L

Table 2: Distribution Line Harmonics and their Impedance for Cases of 09:00Hrs

<table>
<thead>
<tr>
<th>D/L</th>
<th>Length (km)</th>
<th>SV (kV)</th>
<th>RV (kV)</th>
<th>Harmonics</th>
<th>Harmonic Imp. (Ω)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALG-ADM</td>
<td>6.13</td>
<td>33</td>
<td>32.711</td>
<td>2nd, 3rd, 8th, 12th, 16th, 20th, 23rd, 26th, 28th, 29th</td>
<td>10^2.03, 10^2.05, 10^2.28, 10^2.30, 10^2.56, 10^1.90, 10^1.78</td>
</tr>
<tr>
<td>ALG-ALG/L</td>
<td>0.15</td>
<td>33</td>
<td>32.997</td>
<td>2nd</td>
<td>10^2.05</td>
</tr>
<tr>
<td>ALG-ANI</td>
<td>6.84</td>
<td>33</td>
<td>32.645</td>
<td>2nd, 3rd, 6th, 11th, 14th, 18th, 21st, 23rd, 25th, 26th, 27th</td>
<td>10^2.12, 10^2.14, 10^2.06, 10^1.97, 10^1.91, 10^1.84, 10^1.69, 10^1.49, 10^1.99</td>
</tr>
<tr>
<td>ALG-FOW</td>
<td>3.00</td>
<td>33</td>
<td>32.860</td>
<td>2nd, 6th, 13th, 24th</td>
<td>10^1.99, 10^2.27, 10^2.20</td>
</tr>
<tr>
<td>ADM-MAR</td>
<td>1.80</td>
<td>32.71</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ADM-ANI</td>
<td>2.40</td>
<td>32.71</td>
<td>32.645</td>
<td>2nd, 8th, 10th</td>
<td>10^1.98, 10^2.27, 10^2.28</td>
</tr>
<tr>
<td>ALG-BAN/I</td>
<td>5.00</td>
<td>33</td>
<td>32.942</td>
<td>2nd, 3rd, 9th, 15th, 20th, 24th, 29th</td>
<td>10^2.08, 10^2.05, 10^2.01</td>
</tr>
</tbody>
</table>
At 21:00Hrs

**Case 1: ALG – ADM D/L**
Nominal Voltage, V (Vrms) = 117.7kV

![Figure 3i: Harmonics on ALG – ADM D/L](image)

**Case 2: ALG – ANI D/L**
Nominal Voltage, V (Vrms) = 116.63kV

![Figure 3j: Harmonics of the ALG – ANI D/L](image)

**Case 3: ALG – FOW D/L**
Nominal Voltage, V (Vrms) = 91.1kV

![Figure 3k: Harmonics of the ALG – FOW D/L](image)

**Case 4: ADM – ANI D/L**
Nominal Voltage, V (Vrms) = 83.9kV

![Figure 3l: Harmonics of the ALG – FOW D/L](image)

### Table 3: Distribution Line Harmonics and their Impedance for Cases of 21:00Hrs

<table>
<thead>
<tr>
<th>D/L</th>
<th>Length (km)</th>
<th>SV (kV)</th>
<th>RV (kV)</th>
<th>Harmonics</th>
<th>Harmonic Imp. (Ω)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALG-ADM</td>
<td>6.13</td>
<td>33</td>
<td>32.83</td>
<td>2\textsuperscript{nd}, 3\textsuperscript{rd}, 8\textsuperscript{th}, 12\textsuperscript{th}, 16\textsuperscript{th}, 20\textsuperscript{th}, 23\textsuperscript{rd}, 26\textsuperscript{th}, 28\textsuperscript{th}, 29\textsuperscript{th}</td>
<td>(10^{2.01}, 10^{2.64}, 10^{2.27}, 10^{2.15}, 10^{2.05}, 10^{2.01}, 10^{2.93}, 10^{1.86}, 10^{1.74}, 10^{1.56})</td>
</tr>
<tr>
<td>ALG-ALG/L</td>
<td>0.15</td>
<td>33</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALG-ANI</td>
<td>6.84</td>
<td>33</td>
<td>32.73</td>
<td>2\textsuperscript{nd}, 3\textsuperscript{rd}, 7\textsuperscript{th}, 11\textsuperscript{th}, 14\textsuperscript{th}, 18\textsuperscript{th}, 21\textsuperscript{st}, 23\textsuperscript{rd}, 25\textsuperscript{th}, 26\textsuperscript{th}, 27\textsuperscript{th}</td>
<td>(10^{2.12}, 10^{2.66}, 10^{2.37}, 10^{2.14}, 10^{2.06}, 10^{1.97}, 10^{1.91}, 10^{1.84}, 10^{1.69}, 10^{1.49}, 10^{0.99})</td>
</tr>
<tr>
<td>ALG-FOW</td>
<td>3.00</td>
<td>33</td>
<td>32.87</td>
<td>2\textsuperscript{nd}, 6\textsuperscript{th}, 15\textsuperscript{th}, 24\textsuperscript{th}</td>
<td>(10^{2.99}, 10^{2.51}, 10^{2.27}, 10^{2.20})</td>
</tr>
<tr>
<td>ADM-MAR</td>
<td>1.80</td>
<td>32.83</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADM-ANI</td>
<td>2.40</td>
<td>32.83</td>
<td>32.73</td>
<td>2\textsuperscript{nd}, 8\textsuperscript{th}, 19\textsuperscript{th}</td>
<td>(10^{1.98}, 10^{2.47}, 10^{2.28})</td>
</tr>
<tr>
<td>ALG-BAN/I</td>
<td>5.00</td>
<td>33</td>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
At 23:00Hrs

Case 1: ALG – ANI D/L
Amplitude of Source voltage = 32.825kV
Nominal Voltage, V (Vrms) = 116.48kV

Figure 3m: Harmonics of the ALG – ANI D/L

Table 4: Distribution Line Harmonics and their Impedance for Cases of 23:00Hrs

<table>
<thead>
<tr>
<th>D/L</th>
<th>Length (km)</th>
<th>SV (kV)</th>
<th>RV (kV)</th>
<th>Harmonics</th>
<th>Harmonic Imp. (Ω)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALG-ADM</td>
<td>6.13</td>
<td>33</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ALG-ALG/L</td>
<td>0.15</td>
<td>33</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ALG-ANI</td>
<td>6.84</td>
<td>33</td>
<td>32.695</td>
<td>2nd, 3rd, 7th, 11th, 14th, 18th, 21st, 23rd, 25th, 26th, 27th</td>
<td>10^{1.72}, 10^{1.65}, 10^{1.44}, 10^{1.21}, 10^{1.87}, 10^{1.44}, 10^{1.09}, 10^{0.90}</td>
</tr>
<tr>
<td>ALG-FOW</td>
<td>3.00</td>
<td>33</td>
<td>0</td>
<td>-</td>
<td>10^{1.76}, 10^{1.51}, 10^{1.27}, 10^{1.20}</td>
</tr>
<tr>
<td>ADM-MAR</td>
<td>1.80</td>
<td>32.83</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ADM-ANI</td>
<td>2.40</td>
<td>32.83</td>
<td>0</td>
<td>-</td>
<td>10^{1.08}, 10^{1.27}, 10^{1.28}</td>
</tr>
<tr>
<td>ALG-BAN/I</td>
<td>5.00</td>
<td>33</td>
<td>0</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

4. DISCUSSION OF RESULTS

At 02:00Hrs on 17th January 2014, there are 2 Distribution Lines with harmonics namely ALG – ALG/L and ALG – FOW D/Ls. Figures 3a and 3b are harmonic patterns of the Alagbon – Alagbon Local and ALG – FOW D/Ls. For ALG – ALG/L D/L, the 2nd harmonic component is observed on the impedance magnitude versus frequency plotted (harmonic pattern), at a nominal voltage V (Vrms) level of 39.5kV, while the peak impedance magnitude is 10^{1.881}Ω. Whereas, for ALG – FOW D/L, a total of 4 harmonics occurred during simulation and the harmonic impedance values of the harmonics are shown in table 1. The peak impedance magnitude for this case occurs at a value of 10^{1.881}Ω.

At 09:00Hrs, there are harmonics observed on ALG-ADM D/L, ALG-ALG/L D/L, ALG – ANI D/L, ALG – FOW D/L, ADM – ANI D/L and ALG – BAN/I D/L. ALG-ALG/L D/L has only 1 harmonic with peak impedance amplitude of 10^{1.881}Ω and an intercept of 10^{0.40}Ω at nominal voltage of 39.6kV. Furthermore, ALG – FOW D/L has 4 harmonics namely 2nd, 6th, 15th and 24th with peak impedance magnitude 10^{1.76}Ω, 10^{1.51}Ω, 10^{1.27}Ω and 10^{1.20}Ω respectively and all at a nominal voltage of 91.1kV. All other Distribution lines had multiple harmonics, except the ADM – MAR D/L, which apparently, showed no harmonics on the line since the voltage at the source (ADM Bus) was not transmitted to the receiving bus (MAR Bus). A summary is shown in table 2.

At 21:00Hrs, the number of harmonics on ALG – ADM D/L, ALG – ANI, ALG – FOW and ADM – ANI Distribution Lines are respectively, 10, 11, 4 and 3. Their respective Harmonic impedance magnitude values are shown in table 3, while the nature of their harmonic patterns are shown in figures 3 (i,j,k,l).

At 23:00Hrs, Harmonics only occurred in Alagbon – Anifowoshe Distribution Line and about 11 harmonics were observed on the line during this hour as shown in figure 3m. The Harmonic impedance amplitude for this case is summarized in table 4.

5. CONCLUSION

This paper has presented an overview of the Island Business Unit 33kV power Distribution network of Eko Electricity Distribution Plc. An algorithm was developed and used to evaluate the harmonics in each of the distribution lines of the network considered. This was done using MATLAB/Simulink platform. The various harmonics in the distribution lines of the network which were obtained as results are presented. Results were generally discussed and analysed for various scenario times of loading/restoration of 33kV feeders in the network.
REFERENCES


Secured Key Management Technique to Detect Blackhole Attack in Mobile Adhoc Network

O.A. Ojesanmi, O.C. Akinola, O. Oluwasusi & S. Salako

1Department of Computer Science, Federal University of Agriculture, Abeokuta, Nigeria.
2Department of Computer Science, LeadCity University, Ibadan, Nigeria.

ABSTRACT
Mobile Ad-Hoc network (MANET) is very vulnerable to security threats from intruders and attacks that can damage the integrity of the network. This is because of its limited ability to process nodes, bandwidth, and the interaction among the nodes which is involved in decision making within the network. To enhance the strength of security in MANET, we present the use of a routing driven elliptic curve cryptography key to detect blackhole attack in MANET. Simulation will be performed using NetSim 6.0 network simulator on the basis of performance parameters and the effects analyzed after adding Black-hole nodes in the network. The results will be computed and compared to determine its effects on the protocol to achieve better security with integrity.

KEYWORDS - Security, ad-hoc network, MANET, attack, node, routing.

1. INTRODUCTION
A mobile ad hoc network (MANET) is an autonomous system of mobile routers (and associated hosts) connected by wireless links [1][7][15]. The routers are free to move randomly and organize themselves arbitrarily; thus, the network's wireless topology may change rapidly and unpredictably. A routing protocol specifies how routers communicate with each other, disseminating information that enables them to select routes between any two nodes on a computer network, the choice of the route being done by routing algorithms. Routing protocols in MANET can be classified into two categories: reactive protocol and proactive protocol. In proactive routing protocols, all nodes need to maintain a consistent view of the network topology. Reactive routing protocols for mobile ad hoc networks are also called "on-demand" routing protocols. In a reactive routing protocol, routing paths are searched for when needed. Routing protocols are difficult to efficiently secure.

The MANET protocols are facing different routing attacks, such as wormhole replay, colluding misrelay attack, flooding, link spoofing, link witholding and blackhole [2][14]. A large number of attacks have been identified in the recent research that affect the routing in ad hoc wireless networks. Routing attacks can be classified into five categories:

- Attack using a subverted router
- An attack using a rogue router
- Attack using a masquerading router
- Attacks using Modification
- Special Attacks: Other severe types of types of attack are Black hole attack and Wormhole Attack

Secure Routing: These protocols are deals with how a node sends message to other nodes or a base station. A major challenge is to verify authentication of the communication broadcast by the base station. Existing methods often uses public key cryptography which has high computational overhead making them infeasible in wireless sensor network security. The goal of a secure routing protocol is to ensure the integrity, authentication, and availability of messages [10][11]. A comprehensive study of blackhole attacks and countermeasures against this attack in MANET are stated in this paper and the main challenges in focus on designing the robust security solution that can protect MANET from blackhole attacks. This project proposes a newly addition to the Routing security suite with the combination of Secure Ad hoc On-demand Distance Vector (SAODV), Elliptic Curve Cryptography (ECC) Routing protocol with SSH Telnet. The empirical results show that the proposed scheme is quite successful to achieve better solution [3][6][12].

Blackhole Attack: In black hole attack, a malicious node uses its routing protocol in order to advertise itself for having the shortest path to the destination node or to the packet it wants to intercept. This hostile node advertises its availability of fresh routes irrespective of checking its routing table. In this way attacker node will always have the availability in replying to the route request and thus intercept the data packet and retain it. In protocol based on flooding, the malicious node reply will be received by the requesting node before the reception of reply from actual node; hence a malicious and forged route is created. When this route is establish, now it’s up to the node whether to drop all the packets or forward it to the unknown address. The method how malicious node fits in the data routes varies [4][5].
2. MOTIVATION

The next generation security threat seems to be very smart that mobile ad-hoc network needs a sophisticated approach to handle such future threat to make devices on Mobile ad-hoc Network safer. Previous works done on security issues e.g. black hole attack involved in MANET were based on reactive routing protocol like Ad-Hoc On Demand Distance Vector (AODV). Black Hole attack is studied under the AODV routing protocol and its effects are elaborated by stating how this attack disrupt the performance of MANET. Very little attention has been given to study the impact of Black Hole attack in SAODV combine with other protocol and compare the vulnerability of both these protocols against the attack. There is a need to address both these types of protocols as well as the impacts of the attacks on the MANET.

3. METHODOLOGY

The Proposed method will combine Secure Ad hoc On-demand Distance Vector (SAODV), Elliptic Curve Cryptography (ECC) routing protocol with SSH Telnet. Routing protocol uses a hybrid version of AODV protocol. In this proposed protocol, two levels of security are provided. One level is during the route discovery process and the next is during the data transfer. Even if the detection of black hole attack fails at the route discover process, in the next level, it will be identified. So, the proposed protocol has high degree of attack detection and prevention.

Secure AODV: The Secure Ad hoc On-Demand Distance Vector (SAODV) addresses the problem of securing a MANET network. SAODV is an extension of the AODV routing protocol that can be used to protect the route discovery mechanism providing security features like integrity, authentication and non-repudiation. SAODV assumes that each ad hoc node has a signature key pair from a suitable asymmetric cryptosystem. Further, each ad hoc node is capable of securely verifying the association between the address of a given ad hoc node and the public key of that node [8][9].

Telnet: Telnet is a network protocol used to provide a bidirectional interactive text-oriented communications facility, using a virtual terminal connection. User data is interspersed in-band with Telnet control information in an 8-bit byte oriented data connection over the Transmission Control Protocol (TCP). Historically, Telnet provided access to a command-line interface on a remote host. The term telnet may also refer to the software that implements the client part of the protocol. Most network equipment and operating systems with a TCP/IP stack support a Telnet service for remote configuration. Telnet client applications are available for virtually all computer platforms. Telnet means to establish a connection with the Telnet protocol, either with command line client or with a programmatic interface [13].

What is SSH Telnet?

A very reliable security connection solution is SSH or Secure Shell - a network protocol that ensures a high-level encryption, allow data transmitted over insecure networks, such as the Internet, to be kept intact and integrate.

Why use SSH Telnet?: SSH is most often used to establish connections from a certain computer to a remote machine over a network, to execute commands and transfer files between them.

Elliptic Curve Cryptography (ECC) is emerging as an attractive public-key cryptosystem for mobile/wireless environments. Compared to traditional cryptosystems like RSA, ECC offers equivalent security with smaller key sizes, which results in faster computations, lower power consumption, as well as memory and bandwidth savings [3].

3.1 Proposed Security Scheme

SAODV has two types of signatures like single and double. Single Signature (SS) is used for sending RREP from destination and Double Signature (DS) is used for sending route reply (RREP) from intermediate nodes if it has fresh enough route. This work supports both SS and DS. SS is applicable in the initial route discovery process because it is impossible to have enough fresh routes by the intermediate node. If any disturbances occurred during the data transmission process, the source node reinitiates the route discovery process, that time RREP either from destination or from intermediates, both signature process are applicable also the processing steps of SAODV consist of general, intermediate and destination processes.

3.1.1 Proposed Algorithm

The following assumptions are taken in order to design the proposed algorithm.

- Source node broadcasts route request (RREQ)
- Awaits for RREP
- Intermediate node forward request
- Every node has a unique id in the network, which is assigned to a new node collaboratively by existing nodes.
- If intermediate node is blackhole it tunnels the packet to next end
- If packet reaches destination it send reply to source
- On receiving the RREQ it first makes an entry in its Routing table for the node that forwarded the RREQ.
- If it is the Destination node or if it has a fresh enough route to the Destination node, it replies to the RREQ with an RREP.
- Source Node start transmitting data packet through shortest path.
- Send Further Request and Node Identity (NI) of Intermediate Node to Next Hop Node (NHN)
Receive Further Request, NHN of current NHN, Data Routing Information (DRI) entry for
NHN's next hop, DRI entry for current NI
IF (NHN is a reliable node) {
  Check NI for black hole using DRI entry
  IF (NI is not a black hole)
  Then source node send path message to all intermediate node
Intermediate node receive path message
  IF the RREP is received only to the Destination & not to the Restricted IP(RIP), the node carries out the normal functioning by transmitting the data through the route.
  IF the RREP is received for the RIP, it initiates the process of black hole detection, by sending a request to enter into promiscuous mode, to the nodes in an alternate path(i.e. neighbours of next hop for RIP).
  Route data packets (Secure Route)
ELSE {
  INsecure Route
  NI is a black hole
  After receiving RREP from various paths, the source node takes the final decision for choosing the path and that path details are announced by sending SP
  The feedback sent by the alternate paths are analyzed to detect the black hole & this information is propagated throughout the network, leading to the revocation of the Black Holes certificates.

4. SIMULATION

NetSim 6.0 network simulator which is found to achieve the required security with minimal delay and overhead will be used for experimentation. Mobility scenarios are generated using a Random waypoint model by varying 10 to 50 nodes moving in a terrain area of 1500m x 1500m. The simulation parameters are summarized in Table 1. The proposed technique will give a better solution towards Black Hole Attack within the network.

4.1. Simulation Environment

The simulation environment settings are fixed in order to see the fair results between the routing protocols. These routings protocols SAODV and ECC are fairly compared in a chosen free-attack simulation environment.

Here are some of the details on the setup:-

| Table 1: Simulation parameters |
|-----------------------------|----------------|
| Parameter                  | Value          |
| Simulator                  | NetSim         |
| Number of Nodes            | 10 to 50 nodes |
| Maximum connections        | 40 traffic sources |
| Mobility Model             | Random Waypoint |
| Mobility Speed             | 40 m/s         |
| Data Rate                  | 10Mbps         |
| Terrain area               | 1500m x 1500m  |
| Simulation Time            | 500 seconds    |
| Transmission Range         | 250 m          |
| Traffic Model              | CBR            |
| Pause time                 | 2 (s)          |
| Packet Size                | 512 bytes      |
| Routing Protocols          | AODV, SAODV    |
| Pause time                 | 10(s)          |

4.2. Performance Metric

The performance can be evaluated based on the following metrics.

(i) Packet Delivery Ratio.
(ii) Average Throughput.
(iii) Normalized Routing Overhead
(iv) Average Hop Count.
(v) Packet Loss Ratio.
(iv) Average End-to-End Delay.

5. CONCLUSION

In this study, we have proposed techniques to improve the security of ad hoc routing protocols. We proposed the use of SAODV-ECC with SSH Telnet to prevent Blackhole attack. The proposed technique will give a better solution towards Black Hole attack within the network. Black Hole attack with three different scenarios with respect to the performance parameters of Packet Delivery Ratio, Average Throughput and Routing Overhead will be simulated. Future Work: As future work, research work intend to develop simulations to analyze the performance of the proposed solution based on the various security parameters like mean delay time, memory usage, mobility, increasing number of malicious node and increasing number of nodes.
REFERENCES


