IJACSA

WHERE WISDOM SHARES

International Journal of Advanced Computer Science and Applications

Volume 2 Issue 6

June 2011

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

SAI

www.ijacsa.thesai.org

WHERE WISDOM SHARES

INTERNATIONAL JOURNAL OF
ADVANCED COMPUTER SCIENCE AND APPLICATIONS

A Publication of
The Science and Information Organization

IJACSA Editorial

From the Desk of Managing Editor…

It is a pleasure to present our readers with the June 2011 Issue of International Journal of Advanced Computer Science and Applications (IJACSA).

IJACSA is one of the most prominent publications in the field, delivering up-to-date and authoritative coverage of advanced computer science and applications.

The Journal aims to publish the highest quality material, both informative and scientific, on all aspects of Computer Science research. It includes articles related to research findings, technical evaluations, and reviews. In addition it provides a forum for the exchange of information on all aspects.

The Associate Editor Board consists of senior level researchers and/or practitioners, who constantly advise the Co-Editors-in-Chief with respect to the content of and direction to be pursued by the Journal. They may also from time to time be called upon to review manuscripts but that is not their primary role.

Some of the papers have an introductory character, some of them access highly desired extensions for a particular method, and some of them even introduce completely new approaches to computer science research in a very efficient manner. This diversity was strongly desired and should contribute to evoke a picture of this field at large. As a consequence only 32% of the received articles have been finally accepted for publication.

The scope of the journal is fairly broad, although more defined than that of some other journals. Our focus is squarely on experimental research, rather than on work that is largely descriptive.

By having in mind such future issues, we hope to establish a regular outlet for contributions and new findings in the field of Computer science and applications. Therefore, IJACSA in general, could serve as a reliable resource for everybody loosely or tightly attached to this field of science.

And if only a single young researcher is inspired by this issue to contribute in the future to solve some of the problems sketched here or contribute to exiting methodologies and research work, the effort of all the contributors will be rewarded. In that sense we would like to thank all the authors and reviewers that contributed to this issue for their efforts and their collaboration in this project.

We hope to continue exploring the always diverse and often astonishing fields in Advanced Computer Science and Applications.

Thank You for Sharing Wisdom!

Managing Editor
IJACSA
Volume 2 Issue 6, June 2011
editorijacsa@thesai.org
ISSN 2156-5570 (Online)
ISSN 2158-107X (Print)

IJACSA Associate Editors

IJACSA Reviewer Board

CONTENTS

Successful Transmission Rate of Mobile Terminals with Agents in Segmented Ad Hoc Network

Kohei Arai
Department of Information Science
Faculty of Science and Engineering, Saga University
Saga, Japan

Lipur Sugiyanta
Department of Electrical Engineering
Faculty of Engineering, State University of Jakarta
Jakarta, Indonesia

***Abstract*— Mobile Wireless Ad-Hoc Network (MANET) is a special kind of network, where all of the nodes move in time. Random movement is the commonly used to model the mobility in this network. Beside as a host, each node is intended to help relaying packets of neighboring nodes using multi-hop routing mechanism in order to solve problem of dead communication. In recent years, a variety of routing protocols have been proposed specially to accommodate this environment followed with its performance evaluations, likes DSDV, DSR, AODV, and extension OSPF. Researches in this network are mostly simulation based. Research efforts haven't focused much in evaluating network performance when applied to variable number of nodes that involving agents and distributed over several network areas. The paper performed simulations using MANET simulator, which carried out based on the extension OSPF routing protocol. The modification of OSPF protocol had intended to suit the OSPF standard mechanisms for operating with high performance over wireless networks characterized by a broadcast-based transmission medium, in which the physical layer connectivity is not fully meshed. Extensive simulation scenarios have been conducted in the simulator with various numbers of nodes, random and uniform movement direction, and different agent's quantity with different size of network areas. In the performance evaluation of successful transmission (data) packets, the OSPF protocol with throughput weighted metric will be tested under different combination conditions of scenarios.**

Keywords-MANET; agents; movement; packet; routing; network area.

I. INTRODUCTION

Wireless Ad-Hoc Network consists of mobile nodes platforms which are free to move in the area. One class of such networks has been called mobile ad hoc networks (MANETs). Node is referred to a mobile device which equipped with built-in wireless communications devices attached and has capability similar to autonomous router. The most important characteristic of MANET is the dynamic topology, which is a consequence of node mobility. Nodes can change position quite frequently, which means that we need a routing protocol that quickly adapts to topology changes. The node in an ad-hoc network can be laptops and personal digital assistants which are often very limited in resources such as CPU speed, storage capacity, battery power and bandwidth, so the routing protocol should be able to be reactive and capable of minimizing control traffic, such as periodic update messages. To be effective, the routing protocols should have keep the routing table up-to-date and reasonably small, able to choose the best route for certain destination (with weighted metrics in terms of number of hops, reliability, throughput, and cost), and capable of converged within an exchange of a small amount of messages.[8]

In MANET, mobile nodes connected with each other using multi-hop wireless links. Because there is no static infrastructure such as base stations, then if network move arbitrarily thus network topology changes frequently and unpredictably. Each node in the network forwards data packets for other nodes. It makes routing protocol design much difficult. The routing protocol which works well in fixed networks does not show the same performance in MANET. In these networks routing protocols should be more dynamic so that they quickly respond to topological changes [14]. If two nodes are not within proximity of its radio range, all packets must pass through hop by hop by one or more intermediate nodes.

To analyze MANET performance with hundreds or even thousands of nodes, the network must be split into independent layer 3 groups or domains. Smaller domains allow routing, QoS and other network protocols to operate limitary on fewer nodes, with cross-domain interaction only through overlapped area nodes. This division has several key benefits. First, it reduces overall protocol overhead. In most routing protocols, for example, the aggregate route update overhead grows as $O(n^2)$, where n is the number of nodes in a domain. Therefore, using smaller domains with inter-domain interaction through an overlapped area allows reducing the overall overhead. Second, it made network life time longer. Nodes have salient feature of energy-constrained devices. The battery of node is depleted by: (i) computational processing and (ii) transmission/reception of signal to maintain the signal-to-noise ratio above a certain threshold. Although the energy consumption by computations can be further reduced with new developments in low power devices, the energy consumption by communications cannot be overcome. Therefore, partitioning network into smaller domain is essential to develop efficient networking algorithms and protocols that are optimized for energy consumption. As a consequence, when partitioning a network into domains, there are some engineering rules that need to be taken into account. For example, the partition should create balanced overlapped areas; the partitions should minimize the expected traffic among different areas so as to not overload the overlapped area; there must be at least one path between every pair of nodes in certain area and the nodes in overlapped area, and at least one path must exist from overlapped area to nodes that belong to other area for such transmission exchange-area to be successfully take place; there must be at least one node that existed in the overlapped area to act as an relay exchange-area domain. Such nodes are called agents. Third, minimizing and managing exchange-area traffic is desirable since it helps ensuring that the backbone agents do not become congested and maximizes packet forwarding via knowledgeable exchange-area route paths. Agents ensure that no single area suffers adversely from a disproportionately large volume of overhead and data packets.

However, in this paper we focus on the impact of mobility models on the performance of data packets transmission in the large MANET that segmented into two domain areas. So our observations regarding to discuss the effect of movement mobility speed of the nodes to evaluate the performance of the improved OSPF routing protocol with weighted throughput metric, using self-made simulator that considering a dynamic network size with varying number of movement speed at an invariable pause time for mobile Ad-hoc network with frequently and fast moving nodes, movement direction of nodes and domain, either approaching agents or avoiding agents, and different size of overlapped area which potentially vary number of agents getting involved in the traffic at instance of each cycles.

The remainder of the paper is organized as follows. Section 2 gives a brief description of the related work. Section 3 discusses some limitations of the simulation model and provides the simulation model with mobility of our system and describes the effect of some metrics, Section 4 presents the evaluation performance network for data packets transmission, and finally, we conclude the paper in Section 5.

II. RELATED WORK

Significant academic and industrial research has led to the development of a variety of protocols, platforms and architectures for reliable communication provided by the MANET network. They are mostly centralized approaches that assume that reliable communication can be provided by the single uniform network.

Research approached in the field, like SANDMAN [13] and DEAPSpace [12], is done by grouping nodes with similar mobility patterns into clusters; in each cluster, one of the nodes (called cluster head) stays awake permanently and answers discovery requests. The rest of the nodes periodically wake up to provide the actual services and also inform the cluster head about their presence and services [1]. Other approaches take into account the network partitioning and provide related metrics and comparisons are briefly presented in the following paragraphs.

A. *Approach of Expanding MANET to a Large Network*

There are currently no general approaches or methodologies for the creation of domains that take into account the following important engineering constraints: balanced domains, minimal inter-domain traffic, and robust network design [3]. Most existing work on domain generation, however, has used only very limited local information. In fact, most approaches simply elect a "cluster-head" within each subnet based on node attributes like the lowest ID or highest degree [3][4]. Some proposals use local metrics during cluster generation, but the metrics are utilized just for the selection of cluster-heads [5], [6].

Steffano Galli [3] viewed the network as a weighted undirected graph G = (V, E), where each vertex $v \in V$ represents a network node and each edge $e \in E$ represents a viable (potential) communication link with an associated cost we (the cost associated with every edge represents the estimated link load). An example is shown in Figure 1. We wish to partition graph G into N sub-graphs G_i, i =1, 2, …, N, with the objective of satisfying a given set of constraints. It is important to recognize that finding such sub-graphs, i.e. creating network partitions, requires cutting a subset $C \subset E$ of edges in G. The graph G in Figure 1 has been partitioned in N = 4 sub-graphs G1, G2, G3, and G4. The set $C \subset E$ of cut edges is shown in dotted lines.

The rationale for choosing the partition, i.e., for deciding which edges to cut, is based on engineering rules given in the Introduction which can be translated into a graph theoretic framework as follows:

- The partition is as balanced as possible: the number of nodes in every sub-graph G_i, i = 1, 2, …, N, is the same within the smallest possible tolerance.
- The sum of the weights of all cut edges is minimized.
- Each sub-graph G_i, i = 1, 2, …, N, is 1-connected.

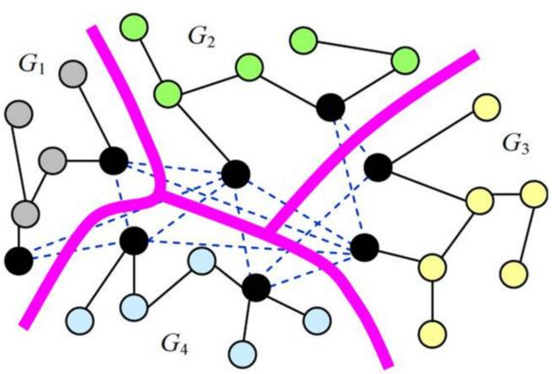


Figure 1. Example of graph partitioning (four areas) of a network [3].

Based on the Galli proposed to solve the OSPF area design problem for large networks using a two-step approach: a first step consisting of an efficient Graph Partitioning Algorithm that will create balanced areas (including a rough Area 0), and a second step consisting of additional ad-hoc heuristics that are able to ameliorate Area 0 design as well as factoring in OSPF specific metrics.

B. *OSPF MANET Extension*

The Open Shortest Path First routing algorithm is widely used for wired networks. As consequences of users' expectation to seamlessly coexist their mobile devices with their wired counterparts, the modification of legacy protocol must be designed. Due to a lack of central base station, nodes in a MANET must form peering relationships to collectively make routing decisions.

OSPF uses a link-state routing protocol to find the least-cost path from a source router to a destination router within a group of routers owned by an organization. As shown in Figure 2, a group of routers using the same routing protocol is collectively referred to as an Autonomous system (AS). Upon joining the AS, a router uses the Hello protocol to discover neighboring routers. Then it forms adjacencies with its new neighbors to exchange routing information; a process called as Link-State Advertisements (LSAs). The exchange (flooding) of routing information, allows routers to obtain complete view of the network topology. Each router stored it as link-state database (LSDB). Initially, LSDB represents the network topology that is visible by router as a result of Dijkstra least-cost path algorithm. In order to maintain and synchronize the network convergence, each router periodically broadcast updates about its current links state to their LSDBs.

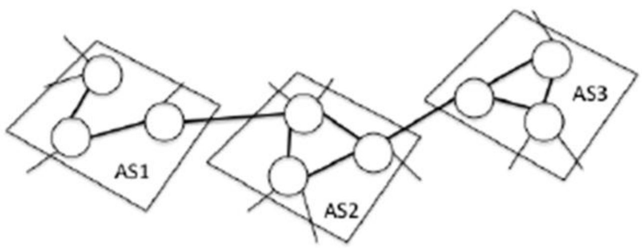


Figure 2. Connected Autonomous systems (ASs)

Although OSPF was designed for wired networks where the topology is predicable and the medium is relatively reliable, its routing algorithm can be extended to support MANET. OSPF is widely deployed and includes many innovative features. The protocol's specifications are in the public domain. It is used by upper-tier Internet Service Providers to determine routes within their networks. It supports multicast routing, multiple same-cost paths, and the ability to organize a network as a hierarchy.

Nodes in a MANET randomly form peering relationships as they move along its communication range. Moreover, packet loss occurs frequently on a wireless medium due to path loss, interferences, noise, shadowing, and multipath. Therefore, OSPF needs to be modified to meet the routing requirements of MANET, which include minimizing data exchange and control overhead.

The OSPFv2 Wireless Interface Type was proposed to support wireless networks. The enhanced protocol is not limited to MANETs; it supports "wireless, broadcast-capable, multi-hop" networks. A team of Boeing Company contributed to the OSPFv2 Wireless Interface Type proposal [4]. Boeing also made other contributions to the OSPF extension for MANETs. They evaluated and analyzed the performance of MANET extensions through simulation of OSPF MANET Designated Routers (MDR) and Cisco's extension (Overlapping Relay and Smart peering) [5][6].

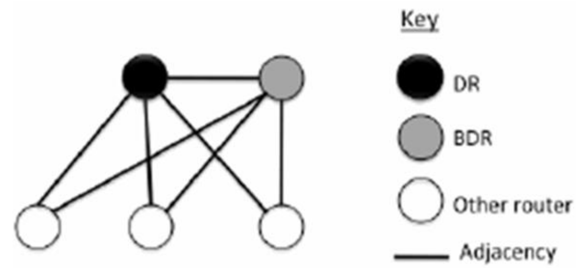


Figure 3. The router's role in the OSPF protocol to reduces the packets adjacencies.

Designated Routers (DR) algorithm provides a mechanism to reduce packets adjacency for data exchanges. The Hello protocol provides a mechanism for electing the network DR. Besides its main function to find and maintain connections among neighbors, Hello packets contain information about router priority that is used to assign the DR. The first router with the highest priority becomes the DR, and the router with the second highest, the Backup Designated Router (BDR) [15]. The DR (router) is responsible to relay the packets to other routers.

OSPF Multi-point Relaying (MPR) uses an MPR selector set to reduce adjacencies. Unlike the OSPF-MDR extension, OSPF-MPR makes not changes to Hello protocol. It minimizes flooding overhead by choosing a flooding MPR set. Similar to Cisco's extension, OSPF Multi-point Relaying (MPR) only forms adjacencies with a limited subset of neighbors to improve scalability and reduce control overhead (LSAs). However, OSPF-MPR does not assign a MDR. A router forms adjacencies with a subset of its one-hop neighbors called MPR selector set.

Most routers only form adjacencies with its MPR selector set, but some Synch routers also form adjacencies with neighbors not in their selector set. A router becomes a synch router only if it has highest router ID among its known neighbors. Hello packets and LSA contain router ID of its neighbors. At least one Synch router must exist in the MANET. However, extra data is added in Hello packets in to facilitate MPR selection.

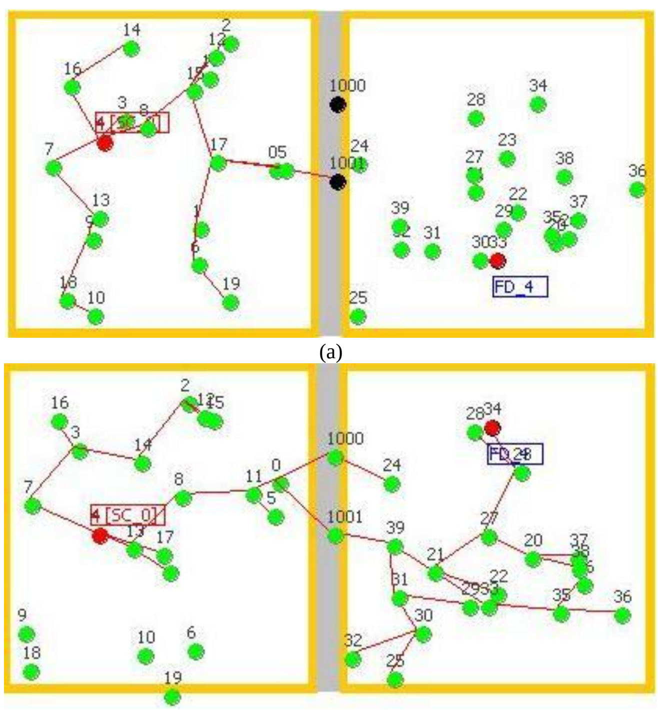
III. SIMULATION MODEL

As foundation for this mobile environment, the core algorithm is developed from static mode (e.g., sensor networks). The enhancement algorithm for serving mobility then detailed in support of topology development, topology maintenance, and routing maintenance.

The model is initiated from broadcast mechanism and propagated through node-to-node based routing metrics approach. Each source injects single big packet which fragmented into multiple packets in the network, which traverse through the network until those reach the final destination. Packets are queued at each node in its path where it waits for an opportunity to be transmitted. This model is not only applicable in direct communication (one hop transmission) but it can also work in multi-hop transmission.

In this situation, when the source and final destination nodes are located outside its clustered network area, source node is capable to discover multiple hop route lead to agents thus maintaining the connectivity required in comparison to standard flooding based ad hoc routing designs.

It is square of Cartesian model area with 200x200 areas and one overlapped section. We consider the case where all nodes in the network are similar, i.e., assuming a homogeneous infrastructure. Inside areas, nodes are deployed uniformly, distributed at random position in the both areas. This deployment produces a connected topology under some assumptions; sometimes a completely connected topology is built and sometimes topology is not fully connected. Simulation build a large connected component quickly using a communications radius considerably smaller than the radius needed to have the entire network connected. Agents are located in such place to facilitate communication between wireless areas and minimize the number of hops to achieve optimum throughput of mesh clients which communicate each other. Agents have static position located in the overlapped between two network areas.



(a)

(b)

Figure 4. Development of network topology within two areas. (a) With Agents involved in delivery of packets. (b) Without Agents involved in delivery of packets.

This mode of messenger can be made clear from the Figure 4. At Figure 4(a), if the source (node 4 mark with red color) coming from certain clustered network area and having destinations (e.g. node 33) located at different (adjacent) clustered network area, the paths are able to meet at a common Agents (node 1000 or node 1001) simultaneously then only any one of the agents will be sufficient to carry all the messages to the proper destination. Thus our routing scheme utilizes the agents capability where a collection of independent request transmission come together for the purpose of cooperative task behaviors and maintaining these connectivity among pairs of sources - destinations. All these tasks essentially work directly through node - agent communication. The entire algorithm works on the fact that agents need to know the existence of each nodes at each clustered areas. On the other side, at Figure 4(b), predefined source (node 4) start the topology development protocol by sending (broadcast) an initial Hello Message. With receive-transmit subsequent routine, the process continues to all reachable nodes. Not every node will be selected to be part of the tree, and those which were not selected will keep silent (in the propagation of packet). Without any knowledge of destination topology (which is located in other network area), then the packets must travel to each vertex to reach the destination node. In both pictures, if there are more than one source nodes starts to transmit packets simultaneously, then several trees may be built in parallel.

A. *The Model*

Simulation consists of multi clustered network environment of homogeneous nodes that communicate with each other using the broadcast services of IEEE 802.11. There are nodes with different roles simulated in this simulation, namely initiator node/source node, receiver node, sender node, destination node, and final destination node. Initiator node/source node is node that initiates transmission of packet. Packet can be either route discovery or data transmission. Like other nodes, initiator is always moving with random direction, speed, and distance. At the time it is moving, initiator node is always sensing its neighbor to maintain connectivity. Receiver node is node that can be reached by source/sender node. Nodes are defined as neighbors if it located within its distance radius range. At initial time, node senses its neighbors before packet data is required to be transmitted. Coverage neighbor nodes always receive packets that are broadcasted from sender. Destination node is selected receiver node in multi hop transmission that should relay packets to the next receiver node. Final destination node is node that became the finish destination of packets.

The layered concept of networking was developed to accommodate changes in local layer protocol mechanism. Each layer is responsible for a different function of the network. It will pass information up and down to the next subsequent layer as data is processed. Among the seven layers in the OSI reference model, the link layer, network layer, and transport layer are 3 main layers of network. The framework is configured in those layers. Genuine packets are initiated at Protocol layer, and then delivered sequentially to next layer as assumed that fragmented packets to be randomly distributed. Simulation models each layer owned with finite buffers. Limited buffer makes packets are queued up according to the drop tail queuing principle. When a node has packets to transmit, they are queued up provide the queue contains less than K elements ($K \geq 1$). To increase the randomization of the simulation process, simulation introduces some delay on some common processes in the network, like message transmission delay, processing delay, time out, etc. This behavior will result that at each instance of a simulation would produce different results. The packets exchanged between sender and receiver is of a fixed rate transmission λ based on a Poisson distribution. Nodes that have packet queued are able to transmit it out using in each available bi-directional link channel.

Our work extends the chance of contacting of the agents that arranged at the fixed cluster from nodes which are distributed randomly within the network. Agents navigating through the network for delivering messages must understand these clustered nodes whenever they are initiating a new route request and thereby increasing the degree of spatial coordination (agents must be on the same place). The temporal coordination has been enhanced with the introduction of a short waiting delay offered to each nodes/agents by the clustered overheads packets. This node – agent coordination will reduce the number of hops and waiting time in spite of further increase the overheads packets hang around with agents and head to highly the agent-chasing problem. The place hosted for the clustered agents can be called as the overlapped area within the network and the detainment period by the agents can be called agent periods.

This overlapped area actually offers a temporary space to be used by all agents for sharing network knowledge, and exchanging messages. Thus when an agent come a fresh it can exploit all other agents who are currently experiencing their agent periods. Here the traveling agents are allowed to carry the information of already visited clusters along with them. The idea behind this is to capture and share the partial network information present with roaming agents. The integration of all such partial information at a common overlapped area helps cooperative tasks like taking the decision for next destination, suitable exchange of messages between agents, getting up-to-date knowledge of the network and reducing unnecessary redundant overhead packets used to visit nodes.

Thus the model of coordination clustered network area where the autonomous agents will be able to deliver messages within a large network with the cooperative communication between them at suitable overlapped area is necessary. There is need of knowing the routes proactively or reactively where part of the network capacity is used for exchanging chunks of routing table data.

We built network simulator to evaluate this proposed algorithm. The simulator supports physical, link and routing layers for single/multi hop ad-hoc networks. We assume that IEEE 802.11 Distributed Coordination Function (DCF) or MAC protocol which uses Channel Sense Multiple Access with Collision Avoidance (CSMA/CA) already deployed. Successfully received packet by receiver's interface is packet whose SNR is above a certain minimum value otherwise the packets cannot be distinguished from background noise/interference. Packets are transmitting through physical layer in accordance with Poisson distribution. Communication between two nodes in IEEE 802.11 uses TCP signaling before the actual data transmission takes place. Simulation simulates this with random hearing to link's condition. If link allow packets to be sent, then sender executes some packets already queued. To execute preferential event in sequentially distributed events, we used a simple approach that consists of applying a different time-event execution by means of the triggering event sequences action. The lower and upper bound of the queuing interval are set such that they do not interfere with predefined timers used by the other events for layers and modification events.

B. Nodes

Nodes are equipped with antenna module installed as capable of dynamically adjusting the transmission energy used to communicate with other nodes. Industrial standard of antenna module supports a management for controlling this energy consumption. The energy consumption required to transmit packet between nodes A and B is similar to that energy required between nodes B and A if and only if the distance and the size of packet are same. The coverage distance range of the nodes is a perfect symmetric unit disk (omni-directional). If $d_{x,y} \leq r_x \rightarrow$ x and y can see each other. This assumption may be acceptable in the condition that interference in both directions is similar in space and time; which is not always the case. Usually interference-free Media Access Control (MAC) protocol such as Channel Sense Multiple Access (CSMA) may exist. In addition, wireless link channel is assumed to have no physical noise; i.e., the errors in packet reception due to fading and other external interferences are not considered as a serious problem. Packets from sender to receiver will be transmitted as long as the bandwidth capacity is sufficient and the received signal to noise ratio (SNR) is above a certain minimum value. Thus every packet successfully received is acknowledged at the link layer and de-encapsulate at the higher layer. Each node is capable of measuring the received SNR by analyzing overheard packet. A constant bit error rate (BER) is defined for the whole network. Whenever a packet is going to be sent, a random number is generated and compared to the packet's CRC. If the random number is greater, the message is received, otherwise it is lost. The default value for the BER is 0, which means there is no packet loss due to physical link error.

Energy is power kept in each node. Heinzelman et al. assumed that the radio dissipates E_{elec} = 50 nJ/bit to run the transmitter or receiver circuitry and ε_{amp} = 100 pJ/bit/m^2 for the transmit amplifier [7]. The radio model is shown in the Figure 5 below.

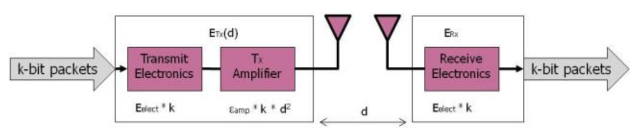


Figure 5. The radio model.

Thus, to transmit a k-bit message a distance d using this radio model, the radio expends:

$$E_{Tx}(k,d) = E_{Tx-elect}(k) + E_{Tx-amp}(k,d) \quad (1)$$

$$E_{Tx}(k,d) = E_{elect} * k + \epsilon_{amp} * k * d^2$$

and to receive this message, the radio expends:

$$E_{Rx}(k) = E_{Rx-elect}(k) \quad (2)$$

$$E_{Rx}(k) = E_{elect} * k$$

Let E_{min_i} is the minimum energy ratio of node I at which a node can still receive, process, and transmit packets. Node j finds out the energy level of neighbor node I through analyzing of received reply packet from node I as it responded the previous transmitted Hello packets. The computation of E_{min_i} is done through two-step propagations. The use of two-steps propagation model is to simulate interactive propagation in the operation of the protocol in dynamic environment. As a future research, the appropriate propagation model that best matches to this environment should replace the simple two-steps model presented here [9][10]. The two-steps propagation model is appropriate for outdoor environments where a line of sight communication existed between the transmitter and receiver nodes and when the antennas are omni-directional. The two-steps propagation model assumes there are two main signal components. The first component is the signal traveling on the line of sight to reached neighbors along with its reply from neighbors and the second component is a confirmation packet transmitted to selected neighbors.

C. Agents

In this paper we introduce mobile agents to hop around the networks for delivering packets .The agents are allowed to meet with other agents at fixed overlapped places as has been mentioned. The cooperation among agents will mutually benefit each other by cooperating in delivering packets. In this current flexible and decentralized framework any autonomous node can send message to any other node at any instant within the network by just issuing a mobile agent. The agent then communicates with other agents to determine the proper one to carries the message to the corresponding destination node. The selected agent then becomes responsible for delivering the message to proper destination. Analog to the real life, these agents actually play the role of messengers and the selected agent play the role of post offices in the ad hoc wireless scenario. Such cooperation among agents scheme has been explicitly designed to reduce the agent traffic in the network. The unnecessary redundant node visits made by the agents to reach destination node has been avoided by sharing and merging with other agents. These agents take the responsibility of providing communication services and improvement of overall traffic in the network.

While delivering messages, an agent will maintain the path records of all the visited nodes in both clustered networks and its corresponding topology. Carrying this network information provide coordination and share the updated network knowledge with other agents. This information field carries topology tree along with numeric values of this membership list (of nodes) collected from each clustered network. In few cases the length of the information list carried by an agent gets longer due to the course of journey made from agent to reach the proper destination. If this happened, the list has been restricted using the hop count limit in order to avoid a huge series of data to be carried along by an agent and subsequent nodes. No packet is allowed to travel forward further whenever its hop count has got exhausted and it is compelled to move back to its originating source node/agent. Thus whenever an agent has finished its forward journey it will eventually follow the same path back to the source node.

The objective of the navigation procedure is to minimize the hops between the agent at overlapped location (current location of the node where the agent is residing) and the source and destination node's location. This criterion would enable an agent to select a neighbor of its current location and take out the packets to the destination nodes. If there is no neighbor available at that instant of time satisfying the above-mentioned criterion, the agent waits for a pre specified amount of time (randomly) and tries to communicate with other agents (any agents can be reached) to get its knowledge. Such contacted agents will respond the request. Through intensive communication among agents, the best agent can be selected to take responsibility for delivering received packets.

In the simulation, when a source node within the ad hoc network wants to send some packets, it immediately senses whether an agent is needed. Each such agent attaches with itself a topology bag to accommodate the request (certain) destination node. This bag is of a given capacity, which can be made full or can be made empty. The source node after initiating the agent puts the packets to (appropriate) agent. The agents are able to exchange these packets with other agents on having suitable coordination with them, which will have the best route path with minimum hops to reach destination node. These agents will be inactive automatically when there is no more packets need to be delivered.

Because of high degree of mobility, the topology will change and it is assumed that the agent will eventually succeed to migrate [2][11]. Whenever an agent wants to leave its current location for delivering packets to some unknown networks it will collect the topology list information from the nodes and will try to reach for a boundary node through which it may get an exit point. The node lying at the boundary will have neighbors from two or more different agents and can act as gateway nodes to other clustered network area. Thus if an agent can reach such a cluster area boundary, it can start visiting a fresh (other) agents. As the location of the agents can be made available from any node of that same clustered network area it can easily track the new nodes there, which has been compulsory. Though the order of cluster visits take place in a random manner still the redundancy in the path visit has been avoided by maintain the path visit list (using BFS function). The agents are free to roam among the overlapped area within the network in a random manner. This agents' capability will be presented on the next paper.

IV. Performance Evaluation

In this section, we evaluate the performance of the routing scheme for packets delivery within clustered ad hoc wireless network areas. We first describe our simulation implementation, performance metrics and methodology, and then we evaluate the agents initiated packets delivery scheme. The results confirm that the agents and their coordination routing algorithm at overlapped area are very efficient in delivering packets under high traffic. The result will also show that living agents in the system as there are packets spread around the clustered network areas; through suitable exchange and message sharing then the number of successful traffics get considerably increased with time.

We started to consider network composed of 60 nodes distributed into two area domain and 2 agents that are located within overlapped area between network areas. The network was set 400x200 areas, with overlapped region width of 20. Fig.6 showed this view of simulation. The movement of the nodes is made random. Nodes are allowed to move at average speeds of 20m/sec. Simulation is run for 100 cycles with movements' parameter was set random (i.e. direction, distance, and speed). Nodes are restricted to move only inside in its area. The following simulation with nodes' movement to other network area, including overlap area will be presented. In the subsequent simulations, the number of nodes is raised up along with combinations of different agents and various width of overlapped area.

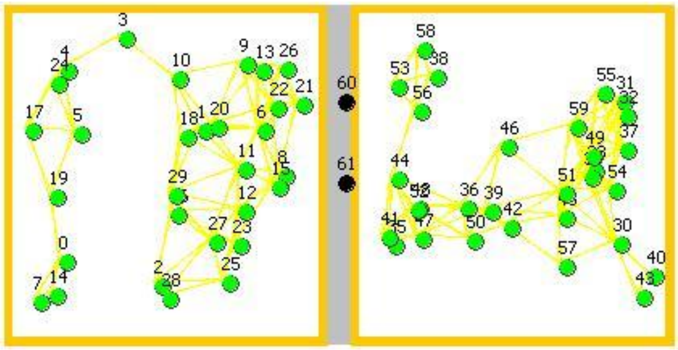


Figure 6. Initial network topology.

The performance of the agents' capability is evaluated using the following two criteria: i) Number of packets whose data contained in the propagation mechanism i.e. the total amount of packet traffic destined at agent to relay packets issued by the nodes in the network. ii) The time period selected as RTT period for successful packets delivery through network; both between source – agent and agent - destination. Network evaluation is intended to investigate the successful transmission rate in the hops between source and final destination nodes and then followed by several network performance analysis that are marked as scenario 1, scenario 2, scenario 3, and scenario 4. At each of scenario, network with similar initiator nodes are loaded with different packets length. We analyze these variables to determine their effects on the minimizing number of hops and average end-to-end delay for packets delivery. The simulation computes different hops/relay connectivity for the packets flow in the delivering packets from source to final destination.

As showed at Fig. 7(a), MANET tends to discover more successful data packets transmission in densely mobile environment (due to moving direction that getting closer to agents). It has been described in section III(b) and III(c), the network topology structure actually consists of two sub protocols, i.e. node topology protocol and agent topology protocol. In the case that many nodes located adjacently each other, thus every node periodically broadcasts Hello packets to denote its presence. Agents do the same sequences to adapt domain environments and to update network topologies for both of proximity areas. This extensive broadcast leads to unavoidable redundant packets for nodes located near overlapped zone. This is depicted in Fig.7(c) where high traffic is identified in the subsequent cycles. These agents, however, are intended to avoid global flooding and save network resources, but in such discovered case, agents are suffered in servicing network.

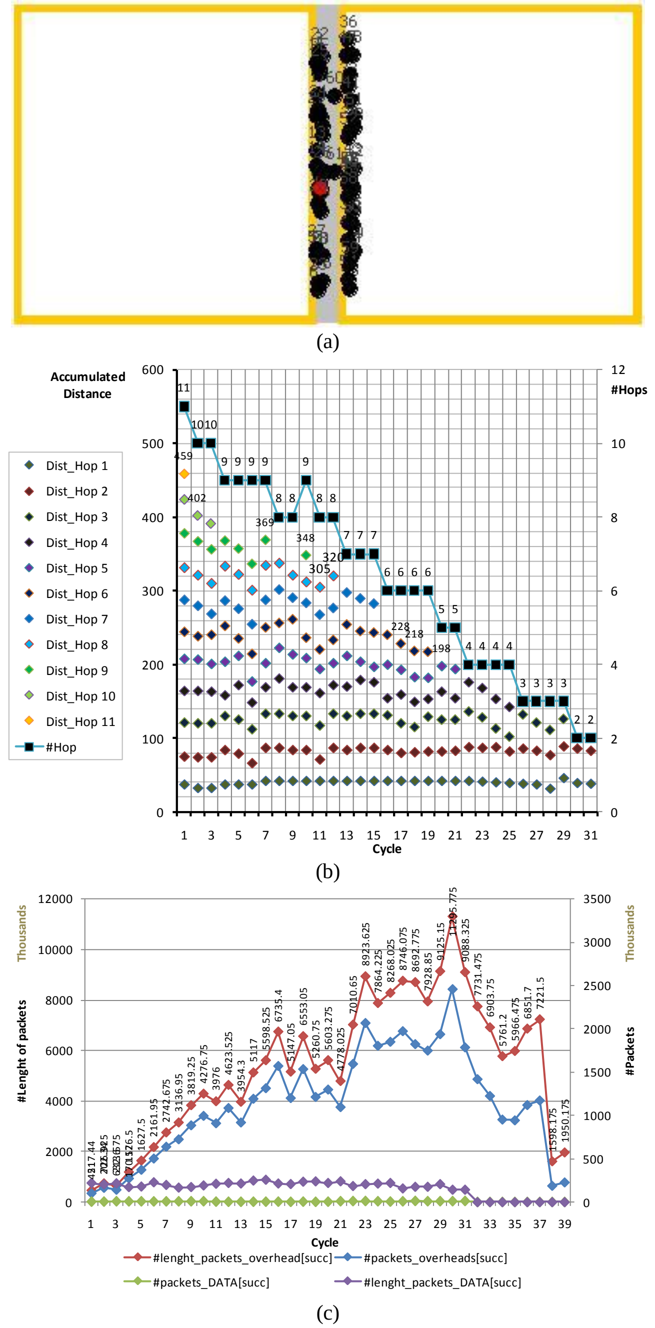


Figure 7. Simulation with nodes' movement direction staticly "closer to Agents" for network with 2 agents and 60 nodes in the two network areas. (a) Network topology at final cycle. (b) Accumulated hops and distances of successful data transmission for different cycle. (c) Broadcasted packets during simulation.

In the highly mobile environments (due to node mobility), more long-successful data packets delivery can be discovered. This is explained by the fact that when the nodes are highly mobile, paths are difficult to be maintained and hence far-away transmission tend to last for a very short amount of time since the probability for a path break is larger when nodes move faster.

When nodes move slower these paths tend to be more stable and hence services tend to be available for a longer time. Such of this occurrence is explained in the Fig. 8.

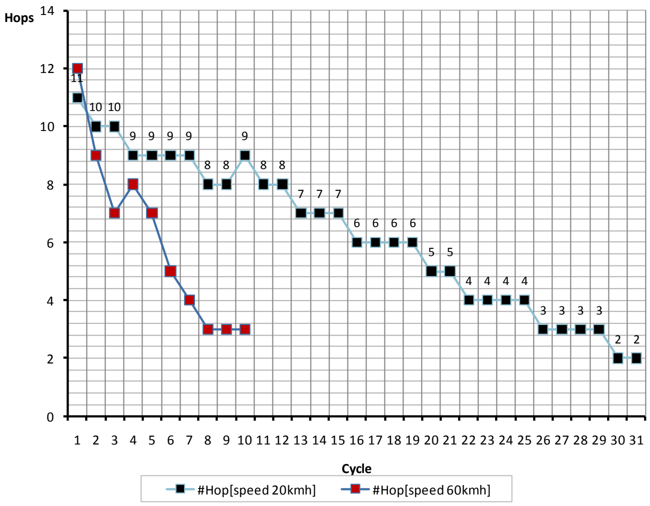
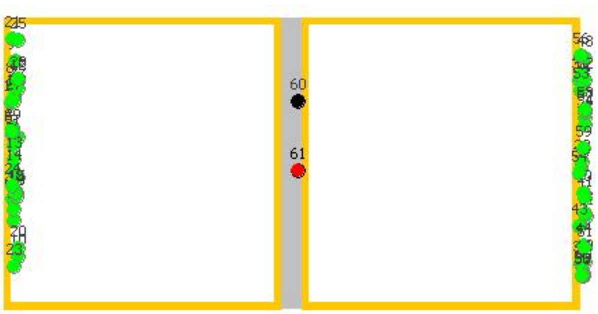


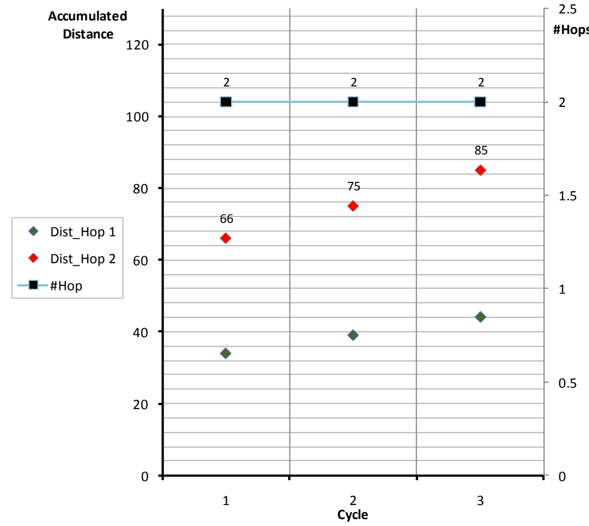
Figure 8. Hops of successful data transmission for different speed of simulation with movement direction "closer to Agents" for network with 2 agents and 60 nodes in the two network areas.

It is obvious from Fig. 7 that maximum successful data packets transmission can be achieved at low mobility when nodes are in overlapped area. The values of average successful data transmission over various mobility speeds are presented in Fig. 8. It is evident from this figure that the average successful data transmission actually decreases when speed increases and its direction are far from agents.

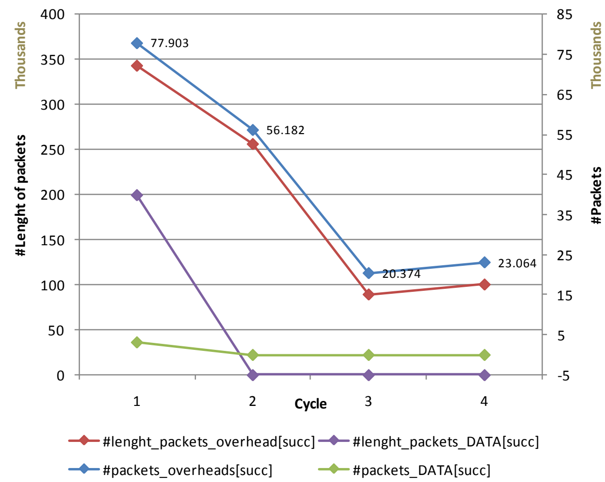
In order to evaluate the movement direction to network traffic, we put aware of the average duration of successful data packet transmission between a node and final destination through various movement directions. As previously predicted, the topology protocol would perform better in low mobility speed and direction of approaching agents. This is explained by the fact that the data transmission services discovered in that setting would be adequate for nodes to complete their transaction. The inverse would be true for low average successful data packet transmission duration, where a setting consists of high mobility speed and direction of away from agents is used for the topology protocol. Such situation can be concluded from Fig. 7, Fig. 8, and Fig. 9.



(a)



(b)



(c)

Figure 9. (a) Network topology at final cycle. (b) Accumulated hops and distances of successful data transmission for different cycle of simulation with movement direction "away to Agents" for network with 2 agents and 60 nodes in the two network areas. (c) Broadcasted packets during simulation.

Simulation was further modified in order to evaluate agents' service information in routing every relayed packet. Each node that is broadcasting this topology update packets sets the TTL (Time to Live) field in these packets equal so that they will be dropped at agents at overlapped area. Every node whose instance location is in overlapped area is set to as agents. Agents listen to information gathered from topology update packets, modify its table and then periodically broadcast its topology table back to nodes in the network. In the situation, where more nodes are located in the overlapped area, overhead traffic among agents leads to significantly higher overhead packets interaction (approximately 125% in average) and at the same time it fail to manage the transmission as no other nodes are discovered nor almost the nodes cannot reach agents. The following figures depict the behavior of network at different movement directions which is grouped as the final situation of network topology (in the (a)), the distance of successful data packets transmission at each hop (in the (b)), and the correlated broadcasted packets during simulation.

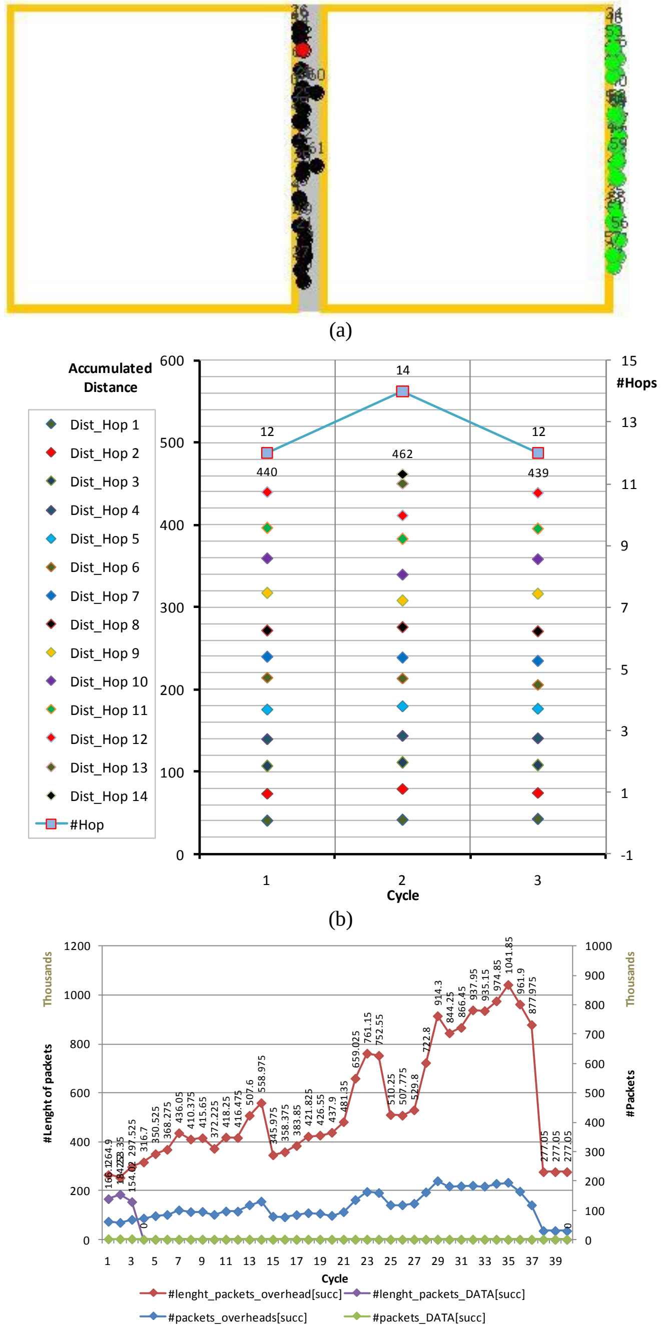


Figure 10. Simulation with nodes' movement direction staticly to East for network with 2 agents and 60 nodes in the two network areas. (a) Network topology at final cycle. (b) Accumulated hops and distances of successful data transmission for different cycle. (c) Broadcasted packets during simulation.

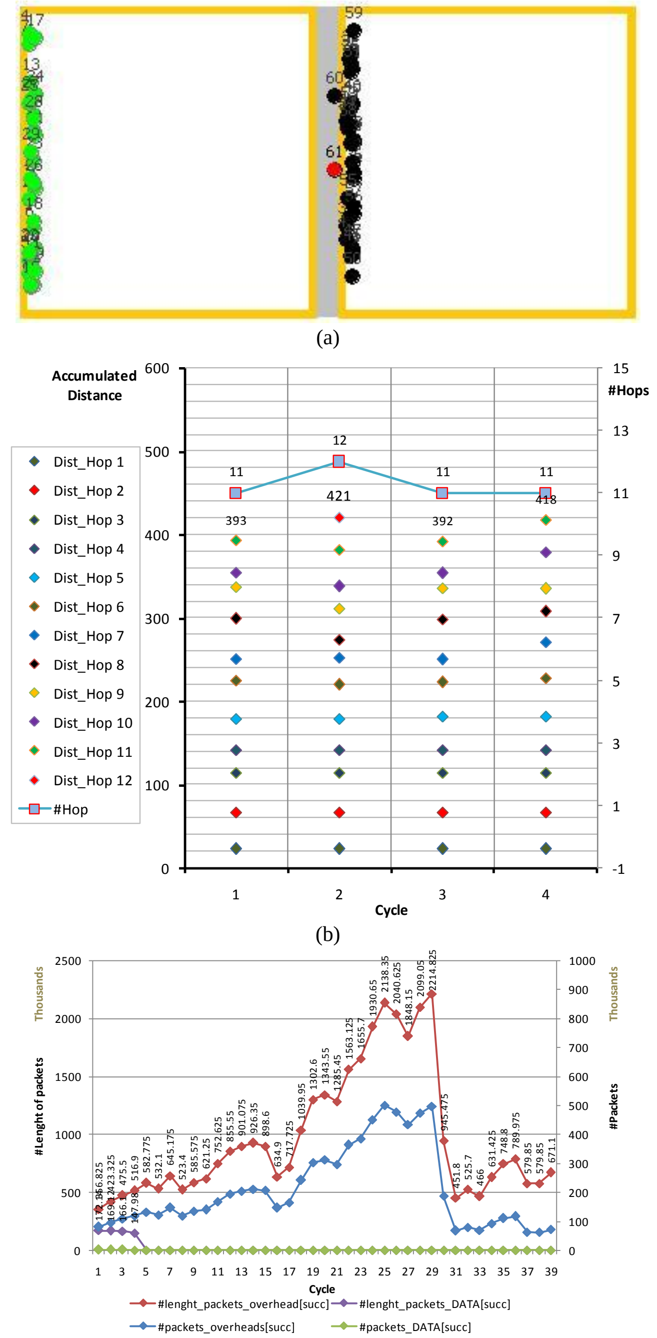
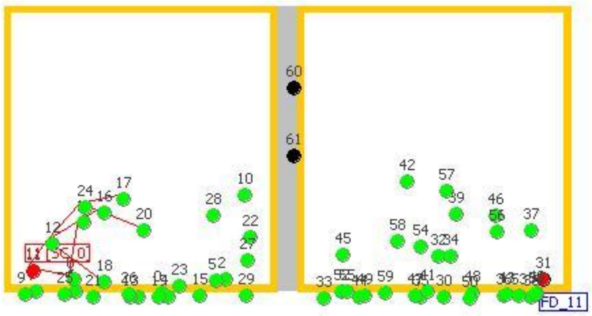
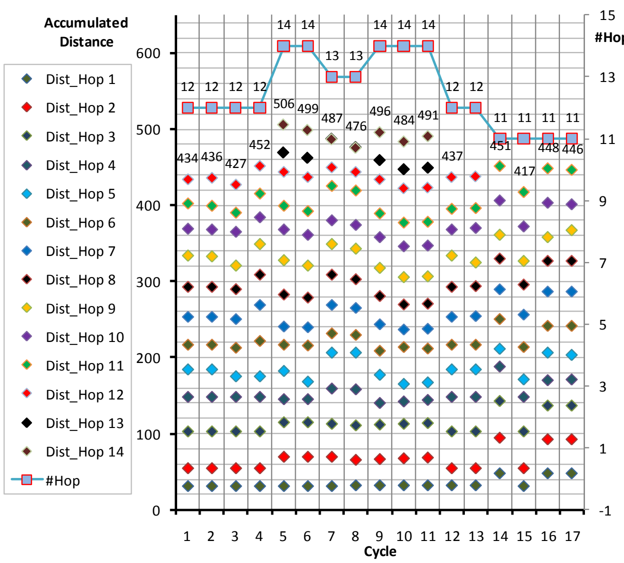


Figure 11. Simulation with nodes' movement direction staticly to West for network with 2 agents and 60 nodes in the two network areas. (a) Network topology at final cycle. (b) Accumulated hops and distances of successful data transmission for different cycle. (c) Broadcasted packets during simulation.

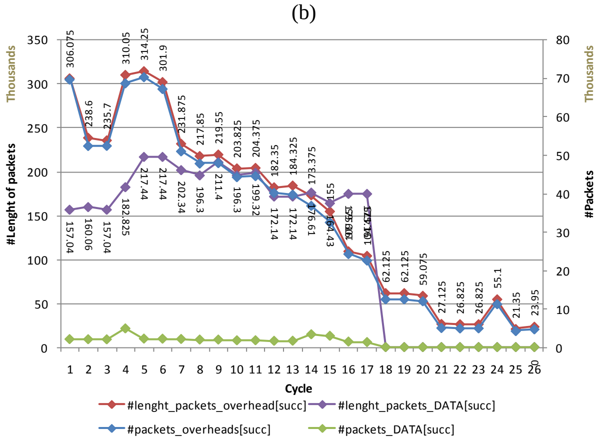
In the Fig. 10(c) and 11(c), with the network movement direction is set to west and east, the broadcasted packets are significantly high at around cycle 28 to 30. During those cycles, almost nodes in one area are located in the overlapped area and acted as agents. Communication among (more) agents and (less) nodes lead to flood the network with overhead packets. After such situation, the traffic tends to decline where nodes and agents are come to a stop at border and overlapped areas.



(a)



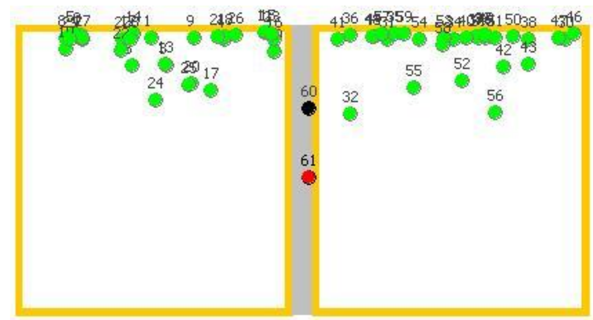
(b)



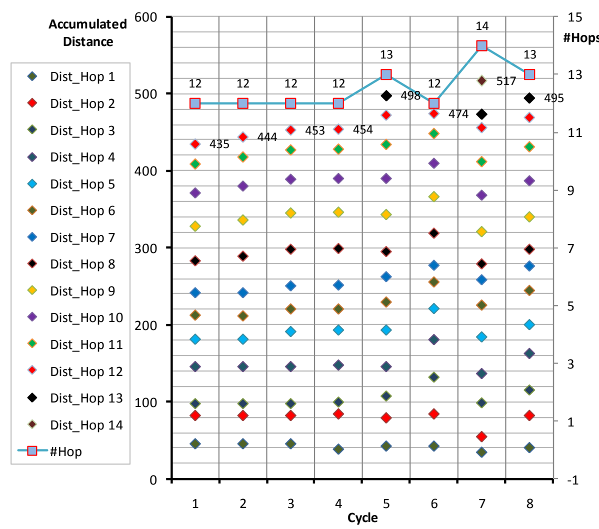
(c)

Figure 12. Simulation with nodes' movement direction staticly to South for network with 2 agents and 60 nodes in the two network areas. (a) Network topology at final cycle. (b) Accumulated hops and distances of successful data transmission for different cycle. (c) Broadcasted packets during simulation.

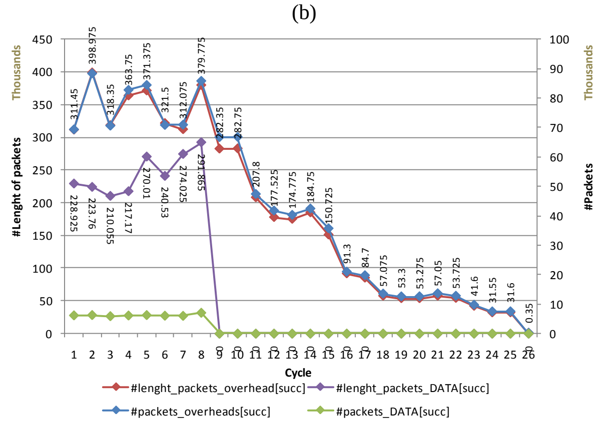
Subsequently, the scenario which every node has speed and direction that is determined of south and north is given below at Fig. 12(c) and the following Fig. 13(c). In all the two simulation models we did this simulation for 50 cycles with 60 nodes, as used previously for Fig. 10 and Fig. 11. Readings were taken for mobility of (speed) 20 km/hour. From the results it is evident that as the nodes coming near the border area increase; the traffic of both figures decline. Low packets nature suggests that more nodes stops looking for agents when no route path founded to reach them in the overlapped area. We simulated static agents in all figures.



(a)



(b)



(c)

Figure 13. Simulation with nodes' movement direction staticly to North for network with 2 agents and 60 nodes in the two network areas. (a) Network topology at final cycle. (b) Accumulated hops and distances of successful data transmission for different cycle. (c) Broadcasted packets during simulation.

Related to density, we simulated two scenarios. The first scenario included 120 nodes that are moving on two areas of total size 400x400 meters, following the random waypoint model with average speed 20m/s (minimum speed is still 0m/s). The second scenario was identical to the previous one but involved only 60 nodes. Both scenarios had duration of 50 step movements each. The results are confirmed with previous scenarios. Different network density showed the same pattern of data packets transmission duration distributions, but the number of successful transmission in the sparsely scenarios is lower on the average of the others found in the first scenario. This is due to the fact that re-discoveries network topology is more frequent in a dense environment. Other fact is also obtained that overhead routing packets are marginally increased in size in order to encapsulate transmission information when density is increasing. The size of overhead packets plays a significant role under high-density cases where congestion is present. Hence, an increase in node density would lead to an increase in overhead packets and such as loss of data packets transmission due to congestion would be discovered.

V. CONCLUSION

In this paper, we evaluate the successful data packets delivery during packets transmission of MANETs with multi-hop routing. With network whose topology always changed all the time, then successful of data packets transmission is mainly depend on the situation of nodes to others. In the densely populated network, this is may not be the issues. On the contrary, if nodes stand apart from other and with their moving directions as such they cannot reach others, then no packets can be transmitted. In the simulation, agents have cascading effects. With their inherent advance capabilities, e.g. relay received packets, to examine the network topologies, determine the next destination of received packet, and guide the route path of packets transmission to reach the destination, agents are mainly to achieve optimum transmission while maintaining connectivity among areas. They can reduce the route of large number of unnecessary packets, particularly in the transmission of packets destined to other area. Except these excellent capability, communication among agents must be take into consideration. Node's movement direction, as key connectivity of nodes to form MANET topology, is evaluated to show the importance of agents. In the dynamic mobile environment, since there are more alternative paths for agents through which node can reach agents and also more alternative agents available in the overlapped area, hence a failure of one or more paths doesn't necessarily mean that the node cannot access the agents in order to deliver packets to reach final destination. Simulation results presented in section IV show that this is not always true, particularly, when nodes are moving away from agents. In such situation, the density increases, despite the existence of multiple paths and agents, the average successful data packets transmission is decreased. This is explained by the fact that pair of source and final destination nodes failed to maintain its connectivity as its distance getting out of agents' proximity. On the contrary, more nodes whose moving direction is getting closer to agents will create more contention for accessing the channel and transmitting service advertisements. Hence, more packet collisions occur and large energy is consumed. The total number of successful data packets transmission however is higher in dense environment with this movement direction. This means that high density may increase the number of successful data packets delivery but it decreases their quality in terms of availability. In nearly future work, we plan to add agents' capabilities to cover the nodes which move through each network areas. Another further investigation need to be conducted on dynamic routing advantages and factors which affect routing mode, e.g., flow type, delay, throughput/delay/reliability tradeoffs between wireless network areas, etc.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for the helpful comments and suggestions.

REFERENCES

[1] Christopher N. Ververidis and George C. Polyzos, "Impact of Node Mobility and Network Density on Service Availability in MANETs", Proc. 5th IST Mobile Summit, Dresden, German, June 2005

[2] Romit RoyChoudhury, Krishna Paul, and S. Bandyopadhyay, "MARP: A Multi-Agent Routing Protocol for Mobile Wireless Ad Hoc Networks" Autonomous Agents and Multi-Agent Systems (Kluwer), 8 (1): 47-68.

[3] Galli et al., "A Novel Approach to OSPF-Area Design for Large Wireless Ad-Hoc Networks", IEEE ICC'05, Seoul, Korea, May 16-20, 2005

[4] Thomas R.H., Phillip A. S., Guangyu P., "Evaluation of OSPF MANET Extensions", Boeing Technical Report: D950-10897-1, The Boeing Company, Seattle, July 21, 2005

[5] Henderson, T., et al., "Evaluation of OSPF MANETExtensions", Boeing Technical Report D950-10915-1, March 3, 2006,

[6] Spagnolo, P., "Comparison of Proposed OSPF MANETExtensions", The Boeing Company, October 23, 2006

[7] Heinzelman, W., Chandrakasan, A., Balakrishnan, H. "Energy-efficient communication protocol for wireless microsensor networks". In: Proceedings of the 33rd International Conference on System Sciences (HICSS): 1–10, 2000.

[8] M. S. Corson, S. Batsell, and J. Macker, "Architecture consideration for mobile mesh networking", Proceedings of the IEEE Military Communications Conference (MILCOM), vol. 1, pp. 225-229, 21-24 Oct. 1996.

[9] Chang-Woo Ahn, Sang-Hwa Chung, Tae-Hun Kim, and Su-Young Kang. "A Node-Disjoint Multipath Routing Protocol Based on AODV in Mobile Adhoc Networks". Proceeding of Seventh International Conference of Information Technology ITNG2010, pp. 828-833, April 2010.

[10] Zheniqiang Ye, Strikanth V. Krishnamurthy and Satish K. Tripathi, "A Framework for Reliable Routing in Mobile Ad HocNetworks", IEEE INFOCOM, 2003.

[11] Romit RoyChouldhury, S. Bandyopadhyay and Krishna Paul, "A Mobile Agent Based Mechanism to Discover Geographical Positions of Nodes in Ad Hoc Wireless Networks", Accepted in the 6th Asia-Pacific Conference on Communications (APCC2000) , Seoul, Korea, 30 Oct. - 2 Nov. 2000.

[12] M. Nidd, "Service Discovery in DEAPspace", IEEE Personal Communications, August 2001, pp. 39-45.

[13] G. Schiele, C. Becker and K. Rothermel, "Energy-Efficient Cluster-based Service Discovery for Ubiquitous Computing," Proc. ACM SIGOPS European Workshop, Leuven, Belgium, Sept. 2004.

[14] N. H. Vaidya, "Mobile Ad hoc networks routing, mac and transport issues", Proceedings of the IEEE International Conference on Computer Communication INFOCOM, 2004.

[15] Cisco, OSPF Design Guide, Document ID: 7039, Aug 10, 2005.

[16] Indukuri, R. K. R. (2011). Dominating Sets and Spanning Tree based Clustering Algorithms for Mobile Ad hoc Networks. *IJACSA -*

International Journal of Advanced Computer Science and Applications, *2*(2). Retrieved from http://ijacsa.thesai.org.

[17] Division, I. S. (2011). A Novel approach for Implementing Security over Vehicular Ad hoc network using Signcryption through Network Grid. *IJACSA - International Journal of Advanced Computer Science and Applications*, *2*(4), 44-48.

AUTHORS PROFILE

Kohei Arai

Prof K. Arai was born in Tokyo, Japan in 1949. Prof K. Arai's major research concern is in the field of human computer interaction, computer vision, optimization theory, pattern recognition, image understanding, modeling and simulation, radiative transfer and remote sensing. Education background:

- BS degree in Electronics Engineering from Nihon University Japan, in March 1972,
- MS degree in Electronics Engineering from Nihon University Japan, in March 1974, and
- PhD degree in Information Science from Nihon University Japan, in June 1982.

He is now Professor at Department of Information Science of Saga University, Adjunct Prof. of the University of Arizona, USA since 1998 and also Vice Chairman of the Commission of ICSU/COSPAR since 2008. Some of his publications are Routing Protocol Based on Minimizing Throughput for Virtual Private Network among Earth Observation Satellite Data Distribution Centers (together with H. Etoh, Journal of Photogrammetory and Remote Sensing Society of Japan, Vol.38, No.1, 11-16, Jan.1998) and The Protocol for Inter-operable for Earth Observation Data Retrievals (together with S.Sobue and O.Ochiai, Journal of Information Processing Society of Japan, Vol.39, No.3, 222-228, Mar.1998).

Prof Arai is a member of Remote Sensing Society of Japan, Japanese Society of Information Processing, etc. He was awarded with, i.e. Kajii Prize from Nihon Telephone and Telegram Public Corporation in 1970, Excellent Paper Award from the Remote Sensing Society of Japan in 1999, and Excellent presentation award from the Visualization Society of Japan in 2009.

Lipur Sugiyanta

Lipur Sugiyanta was born in Indonesia at December 29, 1976. Major field of research is computer network, routing protocol, and information security. Education background:

- Bachelor degree in Electrical Engineering from Gadjah Mada University of Indonesia, in February 2000

- Magister in Computer Science from University of Indonesia, in August 2003.

He is now lecturer in Jakarta State University in Indonesia. Since 2008, he has been taking part as a PhD student in Saga University Japan under supervision of Prof K. Arai.

Assessing 3-D Uncertain System Stability by Using MATLAB Convex Hull Functions

Mohammed Tawfik Hussein
Electrical Engineering Department, Faculty of Engineering
Islamic University of Gaza
Gaza City, Gaza Strip

Abstract— **This paper is dealing with the robust stability of an uncertain three dimensional (3-D) system using existence MATLAB convex hull functions. Hence, the uncertain model of plant will be simulated by INTLAB Toolbox; furthermore, the root loci of the characteristic polynomials of the convex hull are obtained to judge whether the uncertain system is stable or not. A design third order example for uncertain parameters is given to validate the proposed approach.**

Keywords- Algorithm; 3-D convex hull; uncertainty; robust stability; root locus

I. Introduction

Dealing with higher order system can be considered a challenge and a difficult problem, therefore the contribution of this paper is to the utilization of existence built- in MATLAB Convex Hull algorithm and functions to handle such control problems with less time consuming as will be illustrated throughout this research paper.

A. *Motivation and objectives*

This paper is dealing with the robust stability of an interval or uncertain system. Developing an algorithm that checks robust stability of third order uncertain system such systems will be an efficient and helpful tool for control systems engineers.

B. *Literature Review*

The problem of an interval matrices was first presented in 1966 by Ramon E. Moore, who defined an interval number to be an ordered pair of real numbers [a,b], with $a \leq b$ [1- 2].

This research is an extension and contribution of previous publications and ongoing research of the author [3]-[9].

C. *Paper approach*

Three dimension (3-D) convex hull approaches is utilized within MATLAB novel codes that is developed to assess 3-D uncertain system stability, and the algorithm associated is discussed and presented in this paper.

II. Uncertain Systems and Robust Stability

Due to the changes in system parameters due to many reasons, such as aging of main components and environmental changes, this present an uncertain threat to the system, therefore such a system need special type of control system called Robust to grantee the stability to the perturbed parameters. For instances , in recent research robust stability and stabilization of linear switched systems with dwell time[10], as well stability of unfalsified adaptive switching control in noisy environments [11] were discussed.

A. *Robust D-stability*

Letting D(p,q) denote the uncertain denominator polynomial, then the roots of D(p,q) lie in a region D as shown in Fig. 1 , then we can say that the system has a certain robust D-stability properly**.**

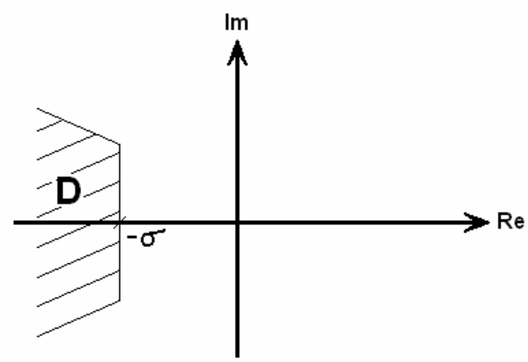


Figure 1. D-region

Definition 1: (D-stability)

Let $D \subseteq C$ and take P(s) to be a fixed polynomial, then P(s) is said to be D-stable if and only if all its roots lie in the region D.

Definition 2: (Robust-D-stability)

A family of polynomials $P=\{p(.,q):q \in Q\}$ is said to be robustly D-stable, if all $q \in Q$, p(.,q) is D-stable, i.e. all roots of p(.,q)lie in D region. For special case when D is the open unit disc, P is said to be robustly schur stable.

B. *Edge Theorem*

A polytope of a polynomial with invariant degree p(s, q) is robustly D-stable if and only if all the polynomials lying along the edges of the polytopic type are D-stable, the edge theorem gives an elegant solution to the problem of determining the root space of polytopic system [12], [13].

It establishes the fundamental property that the root space boundary of a polytypic family polynomial is contained in the root locus evaluated along the exposed edges, so after we generate the set of all segments of polynomials we obtain the root locus for all the segments as a direct location for the edge theorem.

C. Uncertain 3x3 systems

Third order uncertain systems can take the following general form:

$$A = \begin{bmatrix} [\overline{a}_{11}, \underline{a}_{11}] & [\overline{a}_{12}, \underline{a}_{12}] & [\overline{a}_{13}, \underline{a}_{13}] \\ [\overline{a}_{21}, \underline{a}_{21}] & [\overline{a}_{22}, \underline{a}_{22}] & [\overline{a}_{23}, \underline{a}_{23}] \\ [\overline{a}_{31}, \underline{a}_{31}] & [\overline{a}_{32}, \underline{a}_{32}] & [\overline{a}_{33}, \underline{a}_{33}] \end{bmatrix}$$

It has nine (9) elements which mean 2^9 possible combination of matrix family if all elements were uncertain. Generally, we have 2^n possible combinations of an uncertain system where n is number of uncertain elements in the system. Characteristic equations for a general 3x3 matrix can be calculated as shown below in equation (1):

$$P(s) = s^3 - [a_{11} + a_{22} + a_{33}]s^2 + [a_{22}a_{33} - a_{23}a_{32} + a_{11}a_{33} + a_{11}a_{22} - a_{21}a_{12} - a_{31}a_{13}]s + [-a_{11}a_{22}a_{33} + a_{11}a_{23}a_{32} + a_{21}a_{12}a_{33} - a_{21}a_{13}a_{32} - a_{31}a_{12}a_{23} + a_{31}a_{13}a_{22}] \quad (1)$$

The aim of this paper is to calculate the family of all possible combinations for a 3x3 uncertain matrix and so family of possible characteristic equations can be calculated. Then, using convex hull algorithm we will find exposed edges of calculated polynomials and so the roots of exposed edges to determine region of eigenvalues space for studied system.

D. Computing the Convex Hull of the Vertices

The convex-hull problem is one of the most important problems from computational geometry. For a set S of points in space the task is to find the smallest convex polygon containing all points [14].

Definition 1: A set S is convex if whenever two points P and Q are inside S, then the whole line segment PQ is also in S.

Definition 2: A set S is convex if it is exactly equal to the intersection of all the half planes containing it.

Definition 3: The convex hull of a finite point set S = {P} is the smallest 2D polygon Ω that contains S.

III. Methodology And Algorithm

The main goal of this research is to provide a simple and efficient algorithm to determine the bounds of an interval matrix that represent three dimensional problems, hence assess the stability of such an uncertain system by generating a MATLAB Algorithm for three by three interval matrix. Therefore the methodology and algorithm associated with will be discussed and presented in the following sections.

A. Input data and program call

The developed program takes the nine elements of 3x3 uncertain matrix in a vector form and is called in MATLAB command, and these elements can be either real number for specific elements or interval for uncertain matrix entries.

B. Calculating family of possible matrices

To obtain the family of all possible matrices then the following steps are performed within the function **<afamilynew4.m>**:

- Check for size of each input element to find position of uncertain elements.
- Declare input vector of 18 elements containing upper and lower values of elements, **coeff**, if element is specific then upper and lower values are equal.
- Calculate number of uncertain elements in matrix, **ss**.
- Calculate number of possible combinations, $\mathbf{2^{ss}}$.
- Weighting the coefficients vector by use of the function **<weig2.m>** that assign, at each combination, upper or lower values of elements by making use of the idea if binary numbers in combinations. This is done for all the **512** $\mathbf{(2^9)}$ possible combinations.
- Calculate family of **512** possible matrices, **A**.
- Check for repeated matrices and delete it and make sure that remaining matrices are the $\mathbf{2^{ss}}$ unique matrices, **AA**.
- Calculate $\mathbf{2^{ss}}$ possible characteristic polynomials, **polypointsn**, according to equation 2.1. Note that polypointsn is a matrix and is expected to have the size of $\mathbf{(2^{ss},3}$.

C. Find the 3-D convex hull of polynomials

For this purpose, the existence QuickerHull algorithm for convex hulls is utilized and incorporated in the main MATLAB Program [15], [6].

This algorithm has the advantage of being quicker than **convhulln**, the built-in code in MATLAB, as illustrated in Fig. 2.

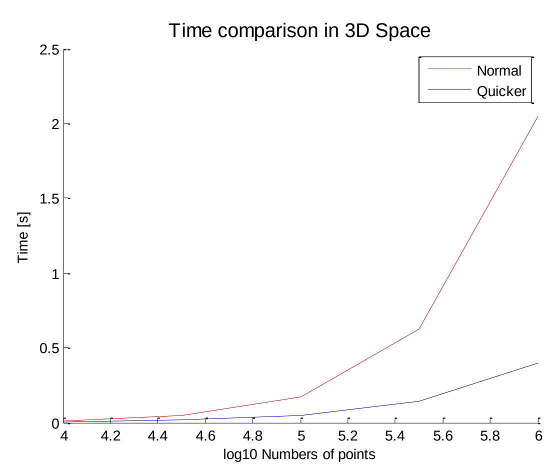


Figure 2. Comparison of processing time in 3D between normal and quicker hull

Then, 3D convex hull of system under study is plotted according to the MATLAB QuickerHull code shown below.

```
function tess=QuickerHull(P)
%  QuickerHull N-D Convex hull.
%    K = QuickerHull(X) returns the indices K of the points in X that
%    comprise the facets of the convex hull of X. X is an m-by-n array
%    representing m points in n-D space. If the convex hull has p facets
%    then K is p-by-n.
%
%    CONVHULLN first attempts to clear points that cannot be part of the
%    convex hullthan uses Qhull. For dimensions higher than six and
%    point’s number less than 1000 no filtering is done.
%
%    Example:
%
%          X=rand (1000, 3);
%          tess=QuickerHull(X);
%    See also convhulln, convhull, qhull, delaunayn, voronoin,
%            tsearchn, dsearchn.
 %erroro check

if nargin>1
   error('only one input supported')
end

[N,dim]=size(P);

if dim>1
   if dim>5 || N<1000
       %run the normal convhull for high dimensions and a few points
        tess=convhulln(P);
         return
   end
```

```
   else
       error('Dimension of points must be>1');
end

%%Filtering points

ncomb=2^dim;%number of combinations among the dimensions
comb=ones(ncomb/2,dim);%preallocate combination
forbregion=zeros(2^dim,1);

%get all combinations using binary logic
for i=2:dim
   c=2^(dim-i);
   comb(:,i)=repmat([ones(c,1);-ones(c,1)],2^(i-2),1);
end
comb=[comb;-comb];%use  combinations simmetry

%for each combination get forbidden region point
```

```
for i=1:ncomb/2
    vect=zeros(N,1);
     for j=1:dim
       vect=vect+P(:,j)*comb(i,j);
     end
     [foo,forbregion(i)]=max(vect);
      [foo,forbregion(i+ncomb/2)]=min(vect);
end

%get the simplyfied forbidden region
%for each dimension get upper and lower limit

deleted=true(N,1);
for i=1:dim

%get combination with positive dimension
    index=comb(:,i)>0;

%upper limit
    simplregion=P(forbregion(index),i);
    upper=min(simplregion);

%lower limit
     simplregion=P(forbregion(~index),i);
     lower=max(simplregion);
     deleted=deleted & P(:,i)<upper & P(:,i)>lower;
 end

%delete points id that cannot be part of the convhull
index=1:N;
index(deleted)=[];

 %Run QuickHull with the survivors
tess=convhulln(P(~deleted,:));
 %reindex
tess=index(tess);
  end
```

D. *Calculate and sort roots of exposed edges of the convex hull*

First we calculate roots of all exposed edges, and plot the convex hull of all roots of possible characteristic equations.

In order to encircle only imaginary components of roots, we need a special sorting algorithm. We sort values according to real and imaginary axes. Sorting may lead into "zero" values in the imaginary matrix if there are real distinct or repeated roots. So, a process of searching for "zeros" in the imaginary matrix is held by replacing any zero vectors in the imaginary by the following one.

IV. PROGRAM OUTPUTS AND RESULTS

For testing research paper’s program, an example of 3x3 uncertain system of a printer belt-drive system is presented and simulated to validate the proposed approach.

A. Design Numerical Simulation Example

The method proposed in this paper will now be demonstrated using printer belt-drive system that is described mathematically by equation (2) and shown below in Fig. 3.

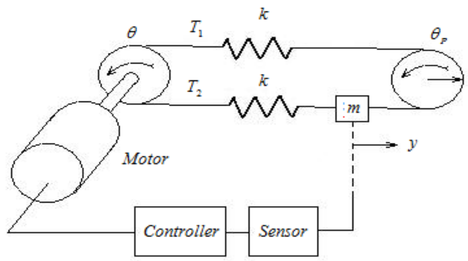


Figure 3. Printer belt-drive system

The system has the following model, see [17]-[19].

$$\dot{x} = \begin{bmatrix} 0 & -1 & r \\ \frac{2k}{m} & 0 & 0 \\ \frac{-2kr}{J} & \frac{-K_m k_1 k_2}{JR} & \frac{-b}{J} \end{bmatrix} x + \begin{bmatrix} 0 \\ 0 \\ \frac{-1}{J} \end{bmatrix} T_d \qquad (2)$$

Where;

r: radius of pulleys [m].
k:spring constant of the belt.
k_1:light sensor constant [V/m].
b: internal friction of the motor [Nms/rad].
R: coil resistance of motor [Ω].
K_m: motor constant [Nm/A].
J: total inertia of motor and pulleys [kgm^2].
T_d: disturbance torque [Nm].

Values of some parameters of this model can vary in the following manner as shown in Table 1:

TABLE I. MODEL PARAMETERS

Parameters	***Values***
m	0.2
k1	1
k2	0.1
r	0.15
b	[0.1 0.25]
R	[1.5 2.5]
Km	2
J	0.01

Then the A matrix is obtained with both lower and upper parameters values as given below:

$$A = \begin{bmatrix} 0 & -1 & 0.15 \\ [10 \quad 400] & 0 & 0 \\ [-1200 \quad -30] & [-13.33 \quad -8] & [-25 \quad -10] \end{bmatrix}$$

B. Source code

With four uncertain elements as shown in matrix A; thus, we expect a family of 16 matrices. By calling the main MATLAB program with the incorporated Quicker Hull functions as presented below:

```
function main4(a11,a12,a13,a21,a22,a23,a31,a32,a33)

polypointsn =
afamilynew4(a11,a12,a13,a21,a22,a23,a31,a32,a33);
size(polypointsn)
tess=QuickerHull(polypointsn)
figure(4)
trisurf(tess,polypointsn(:,1),polypointsn(:,2),polypointsn(:,3))
r=zeros(length(tess),3);
for n=1:length(tess)
   r(n,:)=(roots([1 tess(n,:)]))';
   if imag(r(n,1))~=0
      temp=r(n,3);
      r(n,2:3)=r(n,1:2);r(n,1)=temp;
   end
end
r
x=real(r)
y=imag(r)

figure(5)
plot(x,y,'b+')
hold on;
k=convhull(x,y);
plot(x(k),y(k),'-r')
grid on

q=length(r)
v=0;
z=zeros(1,3);
for i=1:q-1
   if y(i,:)==z
      v=v+1
           z
      for n=i:q-1
         y(n,:)=y(n+1,:);
         x(n,:)=x(n+1,:);
      end
   end
   v1=v;
   y;
   if i==q-1-v
      break
   end
   for s=1:v
      if y(i,:)==z
```

```
        v=v+1
        for n=i:q-1
            y(n,:)=y(n+1,:);
            x(n,:)=x(n+1,:);
        end
      end
   end
end


yn=y(1:q-v,:)
xn=x(1:q-v,:)
v
length(yn)
length(xn)

xxp=xn(:,3);
yyp=yn(:,3);
xxn=xn(:,2);
yyn=yn(:,2);

kp=convhull(xxp,yyp);
kn=convhull(xxn,yyn);
figure(6)
plot(x,y,'b+')
hold on;
plot(xxp(kp),yyp(kp),'-r'), plot(xxn(kn),yyn(kn),'-r')
hold off
grid on

figure(7)
plot(xxp(kp),yyp(kp),'-r'),hold on, plot(xxn(kn),yyn(kn),'-r')
axis equal
grid on
%plot(real(r),imag(r))
%rtess=QuicherHull(r)
end
```

C. Program output

Our proposed program is supposed to show all of family of possible matrices in addition to roots values of characteristic equations.

Four figures are generated while processing our algorithm, Fig. 4 shows 3D convex hull of polynomials. Fig. 5 shows roots locations on s-plane and encircles them by a convex hull. While Fig. 6 shows encircling only imaginary parts of roots, i.e. an identical polygons are generated on and below real axis and other roots are shown also. Fig. 7 focuses on identical polygons encircle imaginary parts of roots.

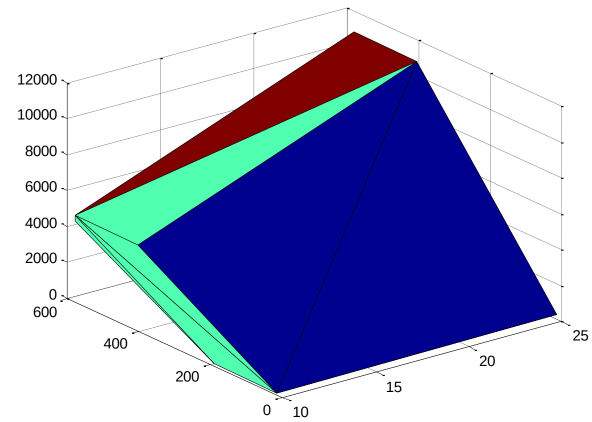


Figure 4. 3D convex hull of system polytopes

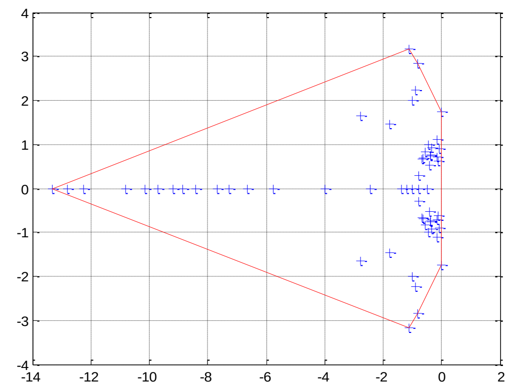


Figure 5. Convex hull of characteristic polynomials roots

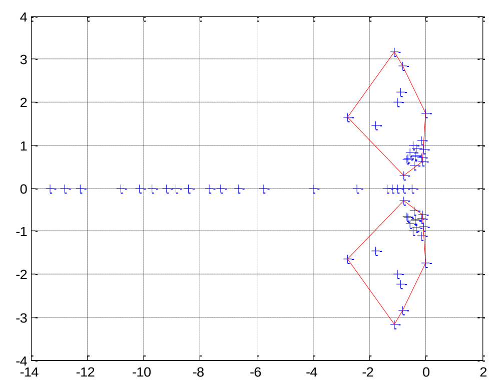


Figure 6. Convex hull of imaginary part of characteristic polynomials roots

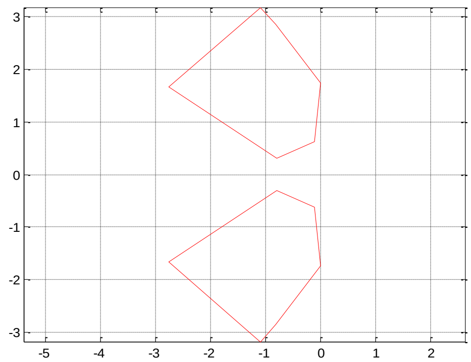


Figure 7. Convex hull of imaginary roots in focus

V. Conclusion And Future Work

In this paper, the stability of uncertain system using convex hull algorithm was tested. And we use MATLAB and INTLAB toolbox to write program that can plot the 3D-convex hull, root loci, step response, and frequency response for any uncertain system. This paper tested the robust stability of an interval 3x3 matrix by the implementation of Printer Belt-Drive System. An efficient and enhanced algorithm was introduced and improved for this purpose.

This algorithm can be easily extended to deal with higher order matrices (n-dimensional system) without a very large increase of processing time.

Acknowledgment

The author wishes to thank his graduate students who enrolled in his course titled “Uncertain Control Systems” during spring 2010 for their inputs, simulations and contribution to this research paper, especially in selecting, testing and running practical engineering examples.

References

[1] R.E. Moore, Interval Analysis. Englewood Cliffs, N.J.: Prentice Hall, Inc., 1966.

[2] L.V. Kolev, “ Interval Mathematics Algorithms for Tolerance Analysis,” IEEE Trans. On Circuits and Systems, 35, pp. 967-975, 1988.

[3] M. T. Hussein," An Efficient Computational Method For Bounds of Eigenvalues of Interval System Using A Convex Hull Algorithm", The Arabian Journal for Science and Engineering, Volume 35, Number 1B, pages 249-263, April 2010,

[4] M.T. Hussein, Lecture Notes for Control of Uncertain Systems Course Faculty of Engineering, IUG, Jan- May 2010.

[5] M.T. Hussein,” Using a Convex Hull Algorithm in Assessing Stability of a DC Motor with Uncertain Parameters: A computational Approach”, IJCSSEIT (ISSN:0974-5807), PP 23-34, Vol. 1, No. 1, January- June 2008.

[6] M.T. Hussein,” A novel Algorithm to Compute all Vertex Matrices of an Interval Matrix: Computational Approach”, International Journal of Computing and Information Sciences, IJCIS (ISSN 1708-0460), 2(2) (2005), pp.137-142.

[7] M.T. Hussein, “A Method to Enhance Tolerance Frequency Analysis of Linear Circuits: Computational Approach”, Al Azher University Engineering Journal, 8(11) (, pp. 373-386, 2005

[8] M. T. Hussein, “A Method for Determination of an Eigenvalue Bounds for an Infinite Family of Interval Matrices: A Computational Approach”, the First International Conference of Science and Development (ICSD-I), March 1-2, 2005.

[9] M. T. Hussein, “Response Analysis of RLC Circuit with an Interval Parameters”, Engineering Week Conference, Gaza, Palestine, December 2004.

[10] L. I. Allerhand and U. Shaked, “Robust Stability and Stabilization of Linear Switched Systems with Dwell Time”, IEEE Trans. on Automatic Control, 56 (2), pp.381-386, 2011.

[11] G. Battistelli, E. Mosca, M. G. Safonov, and P. Tesi, “Stability of Unflasified Adabtive Switching Control in Noisy Envrionments”, IEEE Trans. on Automatic Control,55 (10) ,pp.2424-2429 , 2010.

[12] S.P. Bhattacharyya, H. Chapellat, and L.H. Keel, Robust Control: The Parametric Approach. Prentice Hall, 1995.

[13] S.R. Ross and B.R. Barmish,” A sharp Estimate for the Probability of Stability for Polynomials with Multilinear Uncertainty Structure”, IEEE Trans. on Automatic Control, 53(2), pp.601-606, 2008.

[14] Ketan Mulmuley, Computational Geometry, an Introduction through Randomized Algorithms, Prentice hall, 1994.

[15] INTLAB toolbox User’s Guide Version 3 by math works, http://www.ti3.tu-[harburg.de/rump/intlab/, last visit 20/5/2010.

[16] Luigi Giaccari, http://giaccariluigi.altervista.org/blog/, Created:04/01/2009, last update 08/012009.

[17] C.Bradford Barber, David P. Dobkin, Hannuhuhdanpaa," The Quickhull Algorithm for Convex Hulls", ACM Transactions on Mathematical Software, Vol. 22, No. 4, Pages 469–483, December 1996.

[18] R. C. Dorf, R. H. Bishop, Modern Control Systems, 10th ed., Prentice Hall, Upper Saddle River, NJ 07458, 2005.

[19] K. Ogata, Modern Control Engineering, 4th ed., Prentice Hall, Upper Saddle River, N.J., 2002.

Authors Profile

Dr. Mohammed T. Hussein, an Associate Dean of Engineering for Research and Development and an associate Professor of Electrical Engineering joined the department of Electrical and Computer engineering at Islamic university on August 2003. Dr. Hussein was named Director of e-Learning Center on November 1, 2003. Prior to this appointment he served as a department Head of Engineering Technology in College of Engineering at Prairie View A&M University, Texas. Dr. Hussein earned a Ph.D. degree in electrical engineering from Texas A&M University, College Station, Texas, USA. Dr. Hussein is a registered professional engineer (P.E.) in the State of Texas. Dr. Hussein worked for Motorola Inc., in Tempe, Az., and Oak Ridge National Laboratory in state of Tennessee. His research interests include robust control systems, computer algorithms and applications, and e-Learning. Dr. Hussein holds scientific and professional memberships in IEEE(SM), Eta Kappa Nu, and Tau Beta Pi. He is the recipient of numerous national, state, university, college, and departmental awards including “Who's Who among America’s Best Teachers” on 2000, “Marquis Who’s who among World Leaders” on 2010, and “Teaching Award” in the College of Engineering. Dr. Hussein was nominated and selected on 2003 as an evaluator for Accreditation board for Engineering and Technology (ABET), USA. Dr. Hussein spent summer 2008 as a DAAD visiting Professor at Berlin Technical University, Germany, and on 2009 was selected as academy Fellow, Palestine Academy for Science and Technology.

Automating the Collection of Object Relational Database Metrics

Samer M. Suleiman
Department of Computer Science, Faculty of Computer and Information Technology,
Jordan University of Science and Technology,
Irbid, Jordan

Qasem A. Al-Radaideh, Bilal A. Abulhuda, Izzat M. AlSmadi
Department of Computer Information Systems, Faculty of Information Technology and Computer Sciences,
Yarmouk University,
Irbid, Jordan

***Abstract*— The quality of software systems is the most important factor to consider when designing and using these systems. The quality of the database or the database management system is particularly important as it is the backbone for all types of systems that it holds their data. Many researches argued that software with high quality will lead to an effective and secure system. Software quality can be assessed by using software measurements or metrics. Typically, metrics have several problems such as: having no specific standards, sometimes they are hard to measure, while at the same time they are time and resource consuming. Metrics need also to be continuously updated. A possible solution to some of those problems is to automate the process of gathering and assessing those metrics. In this research the metrics that evaluate the complexity of Object Oriented Relational Database (ORDB) are composed of the object oriented metrics and relational database metrics. This research is based on common theoretical calculations and formulations of ORDB metrics proposed by database experts. A tool is developed that takes the ORDB schema as an input and then collects several database structural metrics. Based on those proposed and gathered metrics, a study is conducted and showed that such metrics' assessment can be very useful in assessing the database complexity.**

Keywords- Object Oriented Relational Database; Metrics; Software Quality.

I. Introduction

The need to store and retrieve data efficiently relative to traditional file structure retrieving was the basic motivation to introduce the relational databases, which is basically the process of representing the data on the form of a collection of related relations (i.e. tables) [1].

Due to the increasing demand on more efficient techniques to store, retrieve and represent complex and huge data types such as images, a new data model is presented with the inspiration of object oriented programming languages. Object Oriented Database (OODB) emerged to meet these demands. OODB is the process of representing the data in a form of complex columns (i.e. objects) that contain attributes and operations to access them [1] [2] [3] [4].

Object Relational Databases (ORDB) have recently evolved for two reasons: the first is the limitation of the traditional relational databases against the increasing demands of the huge applications for storage and fast retrieval of data. The second is the great complexity of the pure OODB [5] [2]. The integration between the relational and the object oriented methodologies could overcome some of the drawbacks that are known in the relational databases, as well as, enable developers to utilize the powerful features of the relational and object oriented databases such as simplicity and usability [5] [6].

Generally, an ORDB system has two main natures: (1) the dynamic nature which reflects the external quality of the system that can be collected from the system at runtime (i.e. dynamic or runtime metrics) (2) the static nature of the system which reflects the internal quality and that can be measured at the design time (i.e. static metrics) [7].

In this scope, metrics are tools to show indications that can help software management in several aspects. For example it can help facilitating the maintenance effort of the schema and hence improve the quality and reduce the complexity of the resulting schema [5]. Controlling the quality of the database system in the design phase may help in preventing the whole system from collapsing in the later phases (e.g. the implementation phase). It also saves the cost and time for the development process in general [8]. The assumption here is that these metrics are standardized and formulated in order to be measured as numbers, and thus facilitates the automation process.

The main metrics for ORDB are: Table Size (TS), Complexity of Weighted Methods (CWM), Cohesion Between Methods (CBM), Coupling Between Objects (CBO), Number of Inherited Properties (NIP), Referential Degree (RD), and Depth in the Relational Tree (DRT) [9] [5] [8] [10].

Several researchers deduced that the value of automation process comes from making the collection and the evaluation process for software and system metrics easier in comparison with manual techniques. In an example in this direction, Stojanovic and El-Emam [11] constructed an object oriented prediction model that can detect the faulty classes based on previous data. They described an open source tool for C++ source code that can calculate the object oriented metrics from the interface specifications at the design phase; these metrics are size, coupling, and inheritance.

There are several challenges facing the evaluation of the metrics. One of the main challenges is the ambiguity in the definitions and formulation of these metrics in addition to the nature of the process of collecting, processing and evaluating those metrics [9] [5] [6]. According to [8] , most of the developed software metrics are programs' oriented and they are not dedicated to database systems.

Al-Ghamdi *et al.* [11] described three tools for the collection; analysis and evaluation of object oriented metrics. The tools are: (1) Brooks and Buells: which contains: a parser, a query engine and a model class hierarchy. (2) The second tool for analyzing C++, and (3) The third tool for gathering OO metrics. In addition, they built their own tool which collects and measures inheritance coupling in object oriented systems. They compared their tool with the three other tools in terms of differences and main features.

Scotto *et al.* [12] mentioned that there is no standardization for software measure and metrics where they suggested using an intermediate abstraction layer to handle the frequent changes on the extraction process for these metrics. They used an automated tool to collect several web metrics. They recommend to separate the two primary activities for any tool that is supposed to measure the metrics, these two activities are the process of extraction and storing the information from the source code, and the second one is the process of analyzing these information and get some conclusion from it.

AL-Shanag and Mustafa [13] proposed and built a tool to facilitate the maintenance and understanding effort for C# source code. The assumption is that software maintenance process can benefit indirectly from software metrics through predicting complexity and possible areas of problems in the software code. Such metrics can help the developer and maintenance software engineers in understanding the source code of programs especially those that have no or little documentation. The proposed tool collects several code elements such as: interfaces, classes, member data and methods from the source code. The tool can also collect the following code metrics: Weighted methods per Class, Depth of inheritance tree, Lines of code, Number of public methods, and Data access metric.

As for this research, the main objective is to build a tool that can collect and evaluate ORDB metrics that will enable designers to calibrate and tune the database schemas to increase usability, maintainability and quality for schemas. The automation of this process becomes essential to overcome the complexity of the evaluation process. Based on the most common definitions of these metrics and units, this research automates the collection of these metrics and realizes them by units scale based on some formal definitions for these metrics. For testing purposes, the proposed automation tool assumes that the input schema is syntactically correct with respect to a standard MS SQL database (particularly MS SQL 2003).

II. Research Background

The challenge in OORD metrics is to find suitable definitions and formulas in order to measure these metrics. These measures are assumed to facilitate the process of controlling the quality of the schema which will enhance the overall performance of the associated information systems [9].

The evaluation of the quality of the database schema must be validated formally. These metrics have to be validated formally through both theoretical and practical approaches. Piattini et al. [8] stated that the concentration is on the practical approaches through developing practical experiments. They developed an experiment to validate the ORDB metrics in order to ensure the benefits of these metrics. They proof the formality and validity of these metrics by repeating the same experiments twice in CRIM center in Canada and University of Castilla-La Mancha in Spain and get similar results. These results showed that Table Size (TS) and Depth in the Relational Tree (DRT) can be used as indicators for the maintainability of database tables.

In ORDB, the table consists of two types of columns: the first is the Standard Column (SC) which is defined as integer or dynamic string data types, and the other one is the Complex Column (CC) which is the User Defined Type (UDT) column. According to this categorization, the metrics also are classified as table related metrics. Those can be applied to a table which includes: TS, DRT, RD, and Percentage of Complex Columns of tables (PCC), Number of Involved Classes (NIC), and Number of shared classes (NSC). Other metrics are applied to schema which includes the DRT, RD, PCC, NIC, and NSC [8].

ORDB schema requires an extra metric due to additional capabilities which come from the object oriented features in order to ensure its internal quality which will be reflected on its external quality in terms of understandability, usability and reliability [9].

A. Metrics of ORDB

Justus and Iyakutti [9] defined and formulated the metrics of ORDB based on three schemas. These metrics are:

- *Table Size (TS):* It represents the summation of the size of both the simple columns (SC) which includes the traditional attributes data types, such as integer and varchar, and the complex columns (CC) which represents the User Defined Types (UDTs). The larger the value of this metric leads to a higher maintenance cost [9] [8]. It is calculated as follows:
$$\boldsymbol{\mu}(\mathbf{TS}) = \sum_{\mathbf{i=0}}^{\mathbf{n}} \mathbf{SC_i} + \sum_{\mathbf{j=0}}^{\mathbf{m}} \mathbf{CC_j} \qquad (1)$$
Where $\boldsymbol{\mu}$ is a metric function, SC represents the Simple Columns and n represents their numbers, CC is complex columns and *m's* are their numbers.
- *Complexity of Weighted Methods (CWM):* It represents the summation of the whole complexities for each weighted method in the table [9] [14]. It is calculated as follows:
$$\boldsymbol{\mu}(\mathbf{CWM}) = \sum_{\mathbf{i=0}}^{\mathbf{m}} \mathbf{C_i} \qquad (2)$$
Where C_i is the complexity of method *i*.
- *Cohesion Between Methods (CBM):* It measures the connectivity between two or more methods and it is measured as the proportion between the similar used attributes in the methods of the class to the total number of attributes. High cohesion is desired which indicates that we are grouping together related methods. Low cohesion

should have a negative impact on maintenance [9] [7] [14] [3]. It is calculated as follows.

$$\mu(\mathbf{A}) = \frac{\mathbf{I_i \cap I_j \cap I_k \cap \ldots I_n}}{\mathbf{V_s}} \qquad (3)$$

Where *I* is the instance of the attributes for the class, and *V* is the total number of variables.

- *Coupling Between Objects (CBO):* It represents the dependency or the connectivity between two methods that exist in two different classes. While high cohesion is desired, high coupling is not as it complicates the design and will complicate maintenance effort, update and reuse [9][7]. It is calculated as follows:

  $$\mathbf{Count_{cp}} = \sum_{\mathbf{i=0}}^{\mathbf{n}} \mathbf{I_i} + \sum_{\mathbf{j=0}}^{\mathbf{m}} \mathbf{Minv_j} \qquad (4)$$

  Where *Min* v_j is the number of methods called and the average number of arguments involved in each invocation.
- *Number of Inherited Properties (NIP):* It represents the number of the properties that have been inherited by the child and thus its value determines high coupling complexity as a trade off from its reuse [9]. It is calculated as follows:

  $$\mu(\mathbf{NIP}) = \sum_{\mathbf{i=0}}^{\mathbf{n}} \mathbf{I_i} + \sum_{\mathbf{j=0}}^{\mathbf{m}} \mathbf{M_j} \qquad (5)$$

  Where M_j= [No. of types * *CWM*] + [No. of methods * *CBO*].
- *Referential Degree (RD):* It represents the number of foreign and reverse refernce keys in the database schema [9] [15] [8]. It is calculated as follows:

  $$\mu(\mathbf{RD}) = \sum_{\mathbf{i=0}}^{\mathbf{n}} \mathbf{FK} + \sum_{\mathbf{j=0}}^{\mathbf{m}} \mathbf{Rvref} \qquad (6)$$

  Where *FK* is the foreign key and *Rvref* is the inverse reference.
- *Depth in the Relational Tree (DRT):* It represents the longest referential path between the tables in or of a database schema [9] [10]. It is calculated as follows:

  $$\mu(\mathbf{DRT}) = \sum_{\mathbf{i=0}}^{\mathbf{n}} \mathbf{d_i} \qquad (7)$$

  Where *d* is the distance between the related tables.

B. Metrics Units

In an attempt to unitize the ORDB metrics, Justus S and K Iyakutti [9] proposed to formulate and calibrate some of these metrics. This is because there is a great need for a standard scale for these ORDB metrics to be correlated with software code metrics. Justus and Iyakutti [9] proposed some units for this purpose based on some experimental studies. The units are:

- Column Complexity (*clm*) unit for the TS metric. Figure 1 shows the relation between TS and the cost in term of *clm* unit.

If the value is in one of the scaled ranges then we can conclude its cost and complexity. For example if the TS value is 50 *clm*, this means that the maintenance table cost is between optimal and low scale.

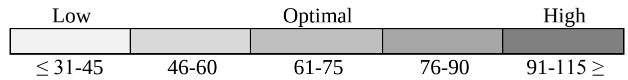


Figure 1: The Table Size unit scale.

- Number of interactions per variables set (*intr/vs*) unit. It is used to measure the cohesion metric as calculated in equation (3).
- Number of messages imported or exported per interaction (*msgs/intr*) unit. It is used to measure the coupling metric as shown in equation (4).

  For both *intr/vs* and *msgs/intr*, Figure 2 shows the relation between these two measures against reusability and maintainability where high *intr/vs* indicates a high class reusability. However, the situation with *msgs/intr* is the opposite which means that when *msgs/intr* is high, this indicates a less class reusability.

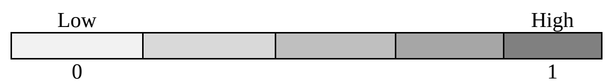


Figure 2: The COM and CBO Units Scale.

- The *laxity* unit is used to measure the reusability metric (NIP). The reusability denotes the usage of the same class-type another time in another class or type. Figure 3 shows the relationship between the reusability of the class-type and this unit. The higher the value of *laxity* the more the probability for the class-type to be reused.

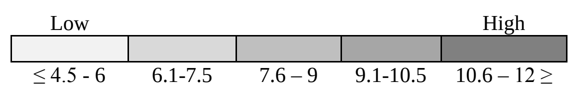


Figure 3: The NIP laxity unit scale.

III. The Developed Tool

The goal of developing this tool is to automatically collect and evaluate ORDB metrics. The proposed tool should enable designers to calibrate and tune the database schemas to increase usability, maintainability and quality for schemas. The proposed tool consists of three main modules: The tokenizer, the lexical analyzer, and the metrics calculator. The lexical analyzer is considered to be the main module of the tool. The architecture of the developed tool is illustrated in Figure 4 and the basic data model for the tool is illustrated in Figure 5.

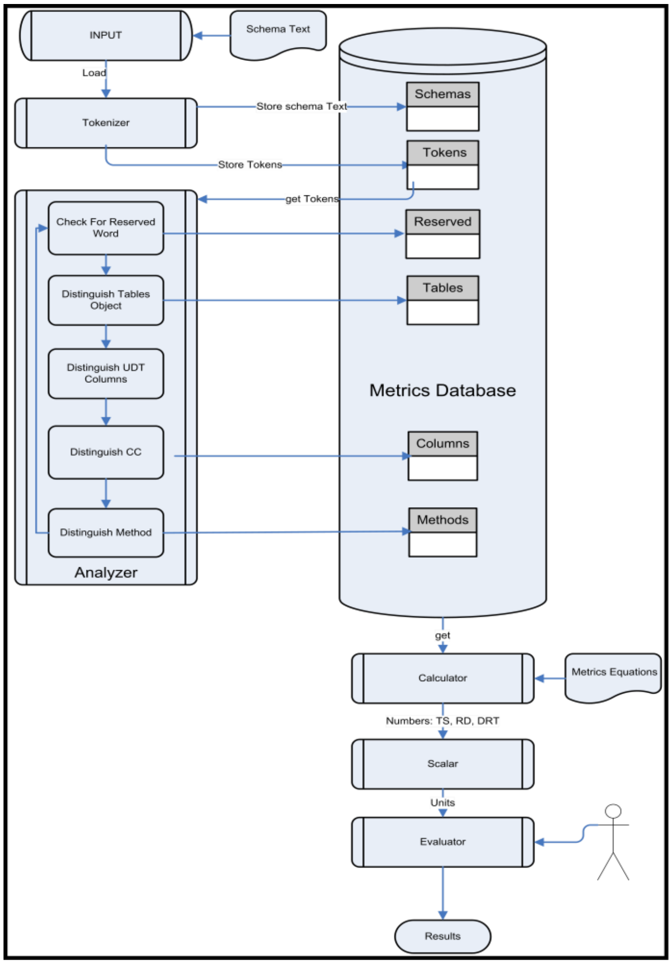


Figure 4: The Lexical Analyzer Architecture

A. *Tokenizer*

The tokenizer aims to facilitate the evaluation process by reading the stream of the text file (i.e. the schema) and tokenize it by recognizing each word individually. In this phase, there is no need to build the relationships between these tokens where the analyzer will handle this job. Tokens will be stored in the database and tagged in order to be used later. Some of these tokens will not be stored in the database for efficiency reasons and because it will not produce any relevant information to the process such as punctuation marks and other language related symbols.

The tokenizer also stores the schema as an entry to the database and links it with other objects and artifacts, in order to create a dataset and accomplish related statistics to obtain a design trend for each dataset. It is worth to mention here that some metrics could be calculated directly such as Lines of Code (LOC). As a summery for this part: the input for the tokenizer is a schema text, and the output is a set of tokens.

B. *Analyzer*

The analyzer is the main part of the tool. It actually accomplishes a great percentage of the evaluation process for the schema. It acts as the bridge between the other parts of the tool. The analyzer starts by reading the tokens from the tokenizer and loads them onto the computer memory in order to be processed and classified into the basic objects for the evaluations. These objects are: the tables, the complex columns and their methods. The challenge here is to reserve the relational structure of the database that will be used subsequently.

The analyzer will check each token against a set of constraints. It will check if the token is an SQL reserved word and if this word is related to the measures or not. For example the table name which can be known if it appears after the two reserved words:"create table" or at least the word, table, hence will be stored in the database and connect all the following distinguished tokens with this entry until the next table name appears.

The tool will store all the information that the metrics' formulas mentioned earlier may need to be collected in efficient and smooth manners. The design of the tool is based on separating the analyzer from the rest of other modules. This may help for any future changes on the metrics equation and guarantee that these changes will not affect the analyses process.

The analyzer distributes the tokens in different connected relational tables and ensures that all the tokens have been stored in the exact place. The equation metrics can easily be applied on these values that we can get from the tables. To summarize: the analyzer inputs are stored tokens, the output from the analyzer categorizes tokens are stored in related tables.

C. *The Calculator Engine*

The calculator engine module uses the information stored in tables to calculate the different metrics presented in section 2. The results are saved in the database associated with each schema for later possible revisions. The calculator module has two sub-modules: the scalar and the evaluator. The scalar role is to map the result of each equation against the scales presented in section II.

The scalar gives meaning to these numbers by assigning a suitable unit to each one of them. The Evaluator evaluates the overall quality of the schema based on the quantized numbers that have been already obtained from the previous units.

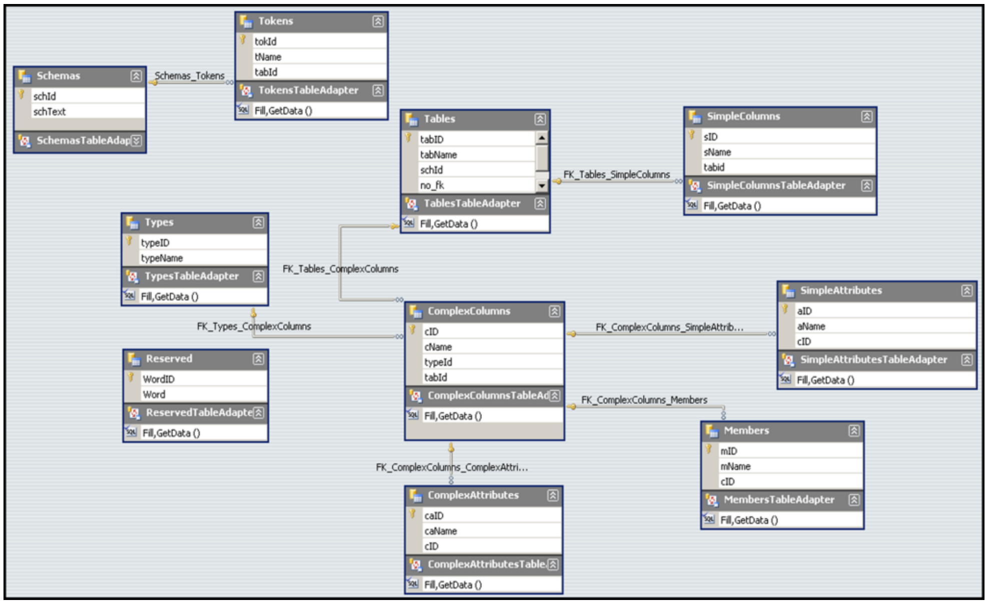


Figure 5: The Tool Data Model

IV. Experiments and Evaluation

In order to evaluate the tool we will provide a complete example that shows how the tool is used to accomplish the tasks. Figure 6 represents a small part of a schema definition adopted from [9] that contains several tables. Table 1 represents the analysis for the table's tokens. Figure 7 shows a sample script for creating a "person_t" type for the same schema definition.

```
CREATE TABLE Tab_Staff
(        emp_no varchar(4),
         person_info person_t,
         do_joining date ,
         working_department department_t ,
         work_for varchar ( 7 ) FK Tab_Student ( roll_no ) inverse ref  ) ;
CREATE TABLE Tab_Student
(        roll_no varchar(7),
         person_info person_t,
         in_department department_t,
        ward_for varchar(4)  );
```

Figure 6: A Sample Schema Definition Script (Justus and Iyakutti, 2008)

As mentioned earlier, the main input for the tool is the schema script (written in SQL 2003 standards) and it is assumed to have object-oriented and relational features that are syntactically correct. The schema is read by the tool as a stream of words or tokens. The main two segments are: the schema definition segment and the implementation segment. They are separated using a special tag or flag.

TABLE 1: Tables of the sample in Figure 6

tabID	tabName	schID	No_fk	No_rev
1	Tab_Staff	1	1	1
2	Tab_Student	1	0	0

```
CREATE TYPE person_t
(   Name varchar(20) ,
    Gender varchar(1) ,
    Birth_date date,
    Address_info address_t,
    MEMBER FUNCTION set_values ( ) RETURN person_t,
    MEMBER PROCEDURE print_person ( ) );
```

Figure 7: A Sample Schema definition for UDT

The tool analyzes this UDT table and relates each member to this new defined type:

- Simple attributes: those represent the standard types just like the simple columns each recognized one is stored in "SimpleAttributes" table and relates to its entry in the CC table. Table 2 represents the record Instances for the "SimpleAttributes" table after the tool reads the script.

TABLE 2. Simple Attributes instances

aID	aName	typeID
1	Name	1
2	Gender	1
3	Birth_date	1

- Complex Attributes: the other possible member items for the UDT is the complex column which is either stored on the "Types" table or need to be treated as we did with the same item member in table type. It can be related to one of the CC entries. Table 3 represents the record Instances for the "ComplexAttributes" table after the tool reads the script.

TABLE 3. Complex Attributes instances

caID	caName	typeID
1	Address_info	1

- Member methods. These members are recognized by the keywords PROCEDURE and FUNCTION, so once the tool has read one of these tokens it will consider the next one as a member function and insert it to the "Members" table and relates it to its entry in the CC table. Table 4 represents the record Instances for the "ComplexAttributes" table after the tool reads the script.

TABLE 4: MEMBERS INSTANCES

mID	mName	typeID
1	set_values	1
2	print_person	1

After this analysis is accomplished, all the necessary information for the schema definition are already stored in the database tables. The following subsection describes the calculation for the Tables Size, RD and DRT metrics, by using the stored data.

A. Calculations

The following sub sections present the calculations process for Table size metrics, Referential Degree, and Depth of Referential Tree

The Calculation of TS

In order to calculate the table size, the following information has to be retrieved from relevant tables:

- "Tables": choose every record which represents a unique table, and retrieve the "tabid", in order to be used to retrieve the related records from the other tables.
- "SimpleColumns": the tool retrieves the related simple columns records from the table for a specific "tabid" value through the "FK_Tables_SimpleColumns" relation (see Figure 8). The number of the retrieved records is stored temporarily in order to be used later.
- "ComplexCoulmns" the tool retrieves the related complex columns records from the table for a specific "tabid" value through the "FK_Tables_ComplexColumns" relation. The number of the retrieved records is stored temporarily in order to be used later.
- The tool checks some other tables which are: "SimpleAttributies", "ComplexAttributies", and "Members". It retrieves all the related records for the type of the complex columns and stores the count sequentially.
- The retrieving process is making use of the following relation between related tables:
  FK_Types_ComplexColumns,
  FK_ComplexColumns_SimpleAttributes,
  FK_ComplexColumns_Members
  FK_ComplexColumns_ComplexAttributes.

The result of this activity is a number that represents the size of the specific type that is mentioned in the table definition. The calculation is based on equation (1) presented in section 2 and the extracted size is stored on the "size" field of the "Types" table. Table 5 illustrates these calculations.

TABLE 5: UDT INSTANCES WITH SIZES

typeID	typeName	Size
1	person_t	14
2	department_t	16
3	address_t	7

By summation of all the retrieved values from these tables, the tool can calculate the size for each table (i.e. number of records from the "SimpleCoulmns" table + size of the existing UDT type(number of records from "SimpleAttributies" table + number of records from "ComplexCoulmns" table + number of records from the "Memebers" table). This is illustrated in Table 6.

TABLE 6: TABLE SIZES

tabID	tabName	schID	No_fk	Size
1	Tab_Staff	1	1	22
2	Tab_Student	1	0	23

To understand these numbers the tool will check them against the proposed scale as in Figure 1. This indicates that tables in this schema fall in the first scale (<= 31-45) and they have low complexity level. Comparing this automatic measurement process with the manual calculations presented by Justus and Iyakutti [9] for TS metric for the same sample schema, the tool shows the same results.

The Calculation of RD metric

The calculation of this metric depends on equation (6) presented in section 2. The tool looks for the existence of the keywords "FK" which stands for foreign key and for "inverse ref" which stands for inverse reference that means the relation is bidirectional between the two subject tables.

The tool counts the frequency for FK token in the schema definition and stores the summation on the "no_fk" field for each table as illustrated in Table 1. The same process is applied on the inverse reference to get the number that represents their counts. In order to calculate the value for this metric, the tool adds the two summations (i.e. sum of FK's and sum of inverse ref's) for each table. Thus we can conclude the complexity of the schema according to the value for this metric. The higher the value of RD means the higher the level of complexity for the schema. Comparing results with [9] for this metric for the same sample schema, the tool gets the same results as they had.

The Calculation of DRT metric

The calculation of this metric depends on equation (7), in which the tool stores the referential path between tables by analyzing the foreign key constraint. It stores the id for the referencing and referenced table. It compares the frequency of the same "tabid" value in both columns and then counts this frequency to get the depth of the referential tree.

The tool gets the length of DRT as a single number that represents the number of tables associated with this relation. Each time it counters the FK keyword, it stores both tables name in the "Tree" table. The first table name after FK keyword is stored in the “DetailTab” field where as the current processing table is stored in “MasterTable” field. Table 7 illustrates these findings:

TABLE 7: DRT FOR THE SAMPLE SCHEMA

MasterTable	DetailTab
Tab_Student	Tab_Staff
Tab_Staff	Exam

The tool then performs an SQL Statement that counts the existence frequency of each table in both columns for each record in the “Tree” table.

B. Tool Evaluation

Table 8 summarizes the results of this research and the output for the developed tool compared with Justus and Iyakutti [9] manual work to calculate the table size. The results in Table 8 show that the tool gets the same results in addition to storing each object in a relational manner in order to be retrieved when they are required. Justus and Iyakutti [9] calculated neither the “Exam” table size nor the “subject-t” type size. Our tool calculates them and compared with the results of manual calculations which are the same.

The same comparison is made between the two approaches with respect to the remaining calculated metrics; the Referential degree (RD) and DRT. The result is illustrated in Table 9, which shows that the automation process facilitates the evaluations for these metrics since the manual approach requires human memorization and tracing.

Figure 8 presents a screen shot of the analysis process that appears for the “Get Tokens” button. All the analyzed tokens appear in the targeted grid view.

V. CONCLUSION AND FUTURE WORK

The evaluation process for the software system quality in general is said to be complex especially for the ORDB metrics. This is due to different aspects such as the lack of formal definitions and standard evaluation. In addition to the lack of such tools that are capable to perform this evaluation.

However, one key point that was investigated in this research is the need to separate the analysis process from the evaluation and calculations processes regarding to these metrics. This separation makes the two main processes independent from one another and handles, partially, the problem of changing the metrics evaluation or definition.

The developed tool performs the analysis process on the schema definition, separates different ORDB artifacts and stores them separately. These artifacts include: tables, their simple columns, their new defined types for the complex columns and the objects created for the complex columns which includes the simple attributes and the members of these objects (i.e. procedures or functions).

The automation process facilitates metrics gathering and evaluation and gives the designers and developers more capabilities and perspectives to ensure the quality of the ORDB systems. It performs the same equations proposed by [9] for TS, DRT, and RD metrics and gets the same results in terms of accuracy compared with manual metrics calculation while improving the performance through calculating those metrics automatically.

The tool is adaptable to the changes of the metrics equations since it separates the analysis process from the evaluation process and this may be useful for the standardization effort by tuning only the evaluation parts of this tool. The analysis may not have to be changed or it may require a little modification. This can help in the continuous evolution of metrics’ formulas’ construction and assessment.

In future, the automation process should be extended to include the remaining proposed metrics such as: (COM) and (CWM). It should be also extended to include more metrics that will be further investigated through looking at different database systems. Another future issue to deal with is that some proposed metrics have no formal scale that enables the designers to conclude the quality of the schema in terms of its complexity and maintainability, and thus future work may define a formal scale for these kinds of metrics.

It is recommended to extend the automation process to include different schemas written in different databases and formats such as: Oracle and MySql. Once these modifications are implemented, there is a need to build a dataset to test them and calibrate their results. The tool may be extended to have the ability to obtain the schema from an existing database and perform the same process as it did with the text schema format.

TABLE 8: COMPARISON BETWEEN MANUAL AND AUTOMATE TABLE SIZE CALCULATIONS

Objects		Type	Manual	The Tool		
Tab_ Staff		Table	23	23=3+14+6		
	working_ department	Department_t	6	Select count(*) from simplecolumns	3	
				person_info, working_department	Person_ info	Working_ department
	person_info	Person_t	14	Select count(*) from simpleAttributes	3	2
				Select count(*) from complexattributes	7	0
				Select count(*) from members;	4	4
	do_joining	Simple Column	1	**Total**	**14**	**6**
	work_for	Simple Column	1			
	emp_no	Simple Column	1			
Tab_ Student		Table	22	22=2+14+6		
	person_info	person_t	14	Select count(*) from simplecolumns	2	
	in_department	department_t	6		person_ info	in_ department
	ward_for	Simple Column	1	Select count(*) from simpleAttributes	3	2
	roll_no	Simple Column	1	Select count(*) from complexattributes	7	0
				Select count(*) from members;	4	4
				Total	**14**	**6**
Exam		Table		13=3+4+6		
				Select count(*) from simplecolumns	3	
					Subject_ info	**for_ department**
	Subject_info	subject_t		Select count(*) from simpleAttributes	4	2
	for_department	department_t		Select count(*) from complexattributes	0	0
	Exam_code	Simple Column		Select count(*) from members;	0	4
	Exam_date	Simple Column		**Total**	**4**	**6**
	Staff_charge	Simple Column				
department _t		UDT	6	6		
	Name	Simple Column	1	Select count(*) from simpleAttributes	2	
	HOD	Simple Column	1	Select count(*) from complexattributes	0	
	set_values	MEMBER	2	Select count(*) from members;	4	
	print_object	MEMBER	2	**Total**	**6**	
Address _t		UDT	7	7		
	Door_no	Simple Column	1	Select count(*) from simpleAttributes	5	
	Area_name	Simple Column	1	Select count(*) from complexattributes	0	
	Zone_name	Simple Column	1	Select count(*) from members;	2	
	City	Simple Column	1	**Total**	**7**	
	Pin_code	Simple Column	1			
	set_values	MEMBER	2			
Person _t		UDT	14	14=3+7+4		
						Address _info
	Name	Simple Column	1	Select count(*) from simpleAttributes	3	5
	Gender	Simple Column	1	Select count(*) from complexattributes	1	0
	Birth_date	Simple Column	1	Select count(*) from members;	4	2
	Address_info	Address_t	7	**Total**	**7**	**7**
	set_values	MEMBER	2		**14**	
	print_person	MEMBER	2			

TABLE 9: A COMPARISON BETWEEN MANUAL AND AUTOMATE RD AND DRT CALCULATIONS

Table Name	Foreign Key Statement	Manual			Automation		
		RD		DRT	RD		DRT
		Foreign Key	inverse ref		Foreign Key	inverse ref	
Tab_Staff	work_forvarchar (7) FK Tab_Student (roll_no) inverse ref	1	1	Memorize master, detail	counterFK + counterINVREF=2		Store Master, Detail
Tab_Student		0	0		0		
Exam	Staff_chargevarchar (4) FK Tab_Staff (emp_no) inverse ref	1	1	Memorize master, detail	counterFK + counterINVREF=2		Store Master, Detail
				Trace for detail for each memorized master			Select count(*) where master=detail from tree=1

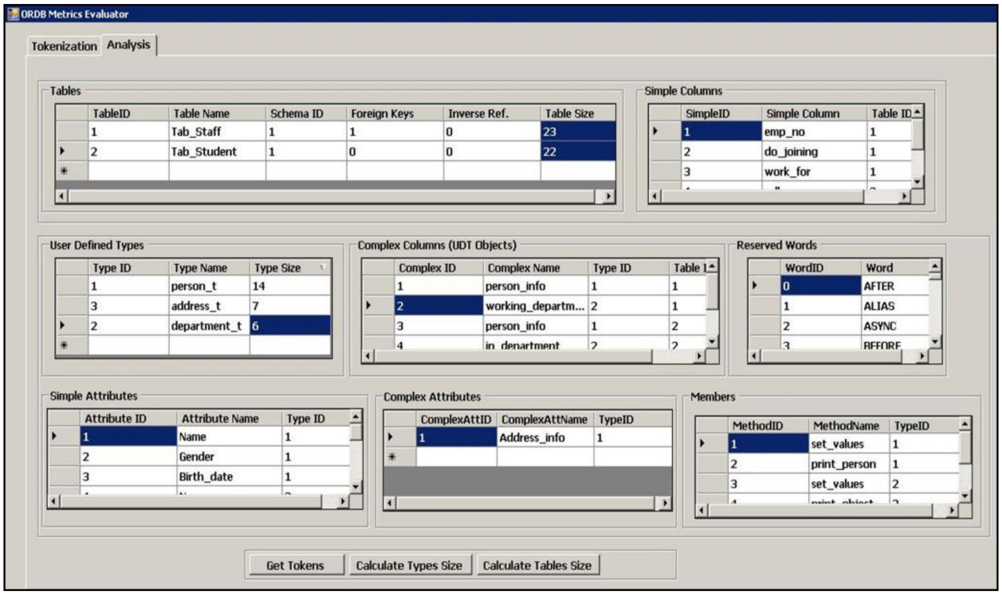


Figure 8: Screen Shot of Results

REFERENCES

[1] Elmasri R. and Navathe S.,"Fundamental of Database Systems", 5th edition, 2007.

[2] Behm A., Geppert A., and Dittrich K.,"On The Migration Of Relational Schemas and Data to Object-Oriented Database Systems" In Proc. of the 5th International Conference on Re-Technologies in Information Systems, Klagenfurt, Austria, December, 1997.

[3] Vaishnavi V., Purao S. and Liegle J. "Object-oriented product metrics: A generic framework", Information Sciences 177 (2007) pp. 587–606.

[4] Van Zyl P. and Boake A., "Comparing the Performance of Object Databases and ORM Tools" in Proceedings of the 2006 annual research conference of SAICSIT**,** 2006**.**

[5] Baroni A.L., Coral C., Brito e Abreu1 F., and Piattini M., "Object-Relational Database Metrics Formalization", Proceedings of the Sixth International Conference on Quality Software (QSIC'06), 2006.

[6] Justus S. and Iyakutti K., "Measurement Units and Scales for Object-Relational Database Metrics", First International Conference on Emerging Trends in Engineering and Technology, 2007(a).

[7] Justus S. and Iyakutti K., "Assessing the Object-level behavioral complexity in Object Relational Databases", International Conference on Software – Science, Technology and Engineering, 2007(b).

[8] Piattini M., Genero M., Coral C. and ALARCOS G., "Data Model Metrics" In Handbook of Software Engineering and Knowledge Engineering: Emerging Technologies, World Scientific, 2002.

[9] Justus S. and Iyakutti K., "A Formal Suite of Object Relational Database Metrics", International Journal of Information Technology 4:4, pp. 251-260, 2008.

[10] Stojanovic M. and El Emam K. "ES1: A Tool for Collecting Object-Oriented Design Metrics", 2001.

[11] Al-Ghamdi J., Elish M. and Ahmed M., "A tool for measuring inheritance coupling in object-oriented systems", Information Sciences—Informatics and Computer Science: An International Journal, v.140 n.3, p.217-227, 2002.

[12] Scotto M., Sillitti A., Succi G., and Vernazz T., "A Relational Approach to Software Metrics", 2004.

[13] Al-Shanag F. and Mustafa S., "Clustering Data Retrieved from C# Source Code To Support Software Maintenance", Master graduation project, Department of computer and information system, Yarmouk university, 2009.

[14] Marinescu R.,"Using Object-Oriented Metrics for Automatic Design Flaws Detection in Large Scale Systems", Object-Oriented Technology (ECOOP'98 Workshop Reader), LNCS 1543, 1998

[15] Mao C., "DBViewer: A Tool for Visualizing and Measuring Dependency Relations between Tables in Database," wcse, vol. 4, pp.13-17, World Congress on Software Engineering, 2009.

A Decision Support Tool for Inferring Further Education Desires of Youth in Sri Lanka

M.F.M Firdhous*
Department of Information Technology
Faculty of Information Technology
University of Moratuwa
Moratuwa
Sri Lanka

Ravindi Jayasundara
Department of Mathematics
Faculty of Engineering
University of Moratuwa
Moratuwa
Sri Lanka

*corresponding author

***Abstract* – This paper presents the results of a study carried out to identify the factors that influence the further education desires of Sri Lankan youth. Statistical modeling has been initially used to infer the desires of the youth and then a decision support tool has been developed based on the statistical model developed. In order to carry out the analysis and the development of the model, data collected as part of the National Youth Survey has been used. The accuracy of the model and the decision support tool has been tested by using a random data sets and the accuracy was found to be well above 80 percent, which is sufficient for any policy related decision making.**

Keywords* – *Educational Desires of Youth; Univariate Analysis; Logit Model; Data Mining.

I. Introduction

Sri Lanka has witnessed several incidents of youth unrest in the recent past. Out of these insurgencies, two insurgencies involved the youth of the south while the other one involved the youth of the north. There have been many discussions and debates about youth unrest and the increasingly violent and intolerant nature of their conflicts. Since these discussions have been rather impressionistic there has always been the need for systematic studies to obtain information on Sri Lankan youth and their background and desires [1]. In order to collect up to date information, targeting to explore facts on Sri Lankan youth and their perceptions, an island wide national youth survey has been carried out. This has been conducted as a joint undertaking involving the United Nations Development programme (UNDP) and other six Sri Lankan and German institutions in the turn of the century. In this survey they have considered four main segments of youth, that is, their politics, conflicts, employment and education. Further Education Desires of youth have been selected to be studied further in this research. The relationship between the types of further education desire of youth in Sri Lanka had been studied with relation to other social factors.

Education domain consists of many different areas but presently in Sri Lanka only a few areas are catered by the national educational institutes [1]. By finding out the educational desires of the youth, it will be possible to design and develop educational and professional programmes and institutes which can be readily accepted by the youth and give better results than that can be achieved by only pursuing traditional programmes. Data Mining which is a powerful tool that can recognize and unearth significant facts, relationships, trends and patterns can be employed to discover this information [2]. In this project, a data mining model has been developed to predict the educational desire of youths at an early stage from other social data.

II. Theoretical Background

Descriptive statistics are used to describe the basic features of the data in a study. They provide simple summaries about the sample and the measures. Together with simple graphics analysis, they form the basis of virtually every quantitative analysis of data [3]. Univariate analysis is the simplest form of quantitative (statistical) analysis.

The analysis is carried out with the description of a single variable and its attributes of the applicable unit of analysis. Univariate analysis contrasts with bivariate analysis – the analysis of two variables simultaneously – or multivariate analysis – the analysis of multiple variables simultaneously. Univariate analysis is also used primarily for descriptive purposes, while bivariate and multivariate analysis are geared more towards explanatory purposes [4]. Univariate analysis is commonly used in the first stages of research, in analyzing the data at hand, before being supplemented by more advance, inferential bivariate or multivariate analysis [5]. Pearson's chi-square test is the best-known of several chi-square tests – statistical procedures whose results are evaluated by reference to the chi-square distribution [6].

With large samples, a chi-square test can be used. However, the significance value it provides is only an approximation, because the sampling distribution of the test statistic that is calculated is only approximately equal to the theoretical chi-squared distribution. The approximation is inadequate when sample sizes are small, or the data are very unequally distributed among the cells of the table, resulting in the cell counts predicted on the null hypothesis (the "expected values") being low. The usual rule of thumb for deciding whether the chi-squared approximation is good enough is that the chi-squared test is not suitable when the expected values in any of the cells of a contingency table are below 5 or below 10 when there is only one degree of freedom [7]. In contrast the Fisher exact test is, as its name states, exact, and it can therefore be used regardless of the sample characteristics [8]. For hand calculations, the test is only feasible in the case of a 2×2 contingency table. However the principle of the test can be extended to the general case of an m×n table [9].

Logistic regression is most frequently employed to model the relationship between a dichotomous (binary) outcome variable and a set of covariants, but with a few modifications it may also be used when the outcome variable is polytomous [10].

The extension of the model and the methods from a binary outcome variable to a polytomous outcome variable can be easily illustrated when the outcome variable has three categories. Further generalization to an outcome variable with more than three categories is more of a notation problem than a conceptual one [11]. Hence, it will be considered only the situation when the outcome variable has three categories.

Main objective of fitting this statistical model is to find out the sequence of variables being significant to the model, so that the sequence of variables, as a whole or a subsequence starting from the first variable, will be used as necessary in constructing a decision tree. In this study we make use of this statistical model not for interpretations but only for doing a comparison with the outcome of a Data Mining approach in decision making.

III. Analysis

Univariate analysis is carried out with the purpose of analyzing each variable independently from other variables. Therefore each of the categorical variables measured as a factor is cross tabularized with the dependent variable "Type of Further Educational Desires" calculating percentages of respondents belonging to each combination of levels, and the Chi-Square Test is performed in order to measure the strength of association between factors and the response of interest. A tolerance rate of 20 percent has been fixed as the significance level for further analysis. Table 1 shows the results of the analysis.

Table 1: Results of Univariate Analysis at 20% Tolerance Level

Explanatory Variable	**Pearson's Chi-square Value**	**Degrees of Freedom**	**Asymptotic Level of Significance**	**Significant Variables at 20% level**
Age group	208.0	6	0.0000	√
Gender	68.751	3	0.0000	√
Educational level	198.1	6	0.0000	√
Ethnicity	53.032	12	0.0000	√
Province	198.1	24	0.0000	√
Sector	6.882	6	0.2020	
Social class (self definition)	77.154	9	0.0000	√
Present financial situation	25.455	6	0.0000	√
Financial situation in past	10.391	6	0.0810	√
Whether school provide good education	7.723	9	0.4860	
Major problems with education	95.518	21	0.0000	√
Access to educational facilities	90.419	6	0.0000	√
Type of activity	1066.0	18	0.0000	√

From the results shown in Table 1, the variables with their p-values less than 0.2 have been detected as significant.

A. *Fitting a Statistical Model*

The main purpose of this part of analysis is to determine the factors, which affect or are associated with having different types of Further Educational Desires in youth. Though several variables have been identified as significant factors, where each could independently build a significant effect on developing different wishes on education among youth, due to the confounding nature of these factors, it is not easy to conclude on their corporative influence on making Further Education Desires different in people.

Therefore this modeling approach can be very much useful in detecting the genuine effect of these factors when adjusted for some other factors as well.

Since the response variable is Multinomial and the scale of response levels are Nominal, it was decided to work out a "logit" link in regression modeling. Therefore a Generalized Logit Model will be fitted to accomplish the objective.

B. *Fitting the best fitted Generalized Logit Model*

The Forward Selection procedure is used in selecting variables to the model. In assessing the fit of the terms to the model, the difference in deviance of the two models compared, which is distributed as Chi-Squared has been used at the 5% significance level. However the terms will be selected to the model, as they do the best representation of all the data.

The results obtained in following the steps of fitting a Generalized Logit Model using the procedure CATMOD in SAS package, are tabularized in the body of the analysis.

Let the Null Model

be, $\mathbf{log}(\frac{\pi_f}{\pi_F}) = \boldsymbol{\alpha_f}$;

where $\boldsymbol{\pi_f}$ is the probability that a respondent has the $\boldsymbol{f^{th}}$ type of Further Education Desire, f $\neq$ **F**, and type F is the Further Education Desire category "No Desire".

<u>Fitting Main Effects to the Model</u>

Step 1: Null Model vs One Variable Model (Model 1)

Table 2 shows sample of data used in devising Model 1.

Table 2: Adding the 1st Most Significant Variable to the Model

Model 1 (terms added)	*Raw Deviance (-2Log likelihood)*	*Difference in Deviance*	*Difference in df*	*p-value*
Null Model	4540.468	0	-	-
Age group	4332.934	207.5343	4	9.00999E-44
Gender	4474.573	65.8957	2	4.90829E-15
Educational level	4359.197	181.2715	4	3.97611E-38
Ethnicity	4490.291	50.1769	7	1.33346E-08
Province	4401.342	139.1265	16	1.06808E-21
Social class	4477.97	62.4984	6	1.39659E-11
Present financial situation	4517.147	23.3216	4	0.000109204
Financial situation in past	4532.136	8.332	4	0.080146321
Major problems with education	4452.271	88.1975	14	8.30242E-13
Access to educational facilities	4450.442	90.0261	4	1.30006E-18
Type of activity	**3450.942**	**1089.527**	**10**	**9.5521E-228**

From Table 2, the lowest p-value is associated with the variable 'Type of activity'. The selection procedure of the most significant variable requires that 'Type of activity' be added to the Null Model as the first step of developing a model where the 'Type of Further Education Desire' being the response variable.

<u>The explanatory variables in the model</u>: Type of activity

Model 1: $\mathbf{log}(\frac{\pi_{if}}{\pi_{iF}}) = \boldsymbol{\alpha}_\mathbf{f} + \mathbf{x}_\mathbf{i}'\boldsymbol{\beta}_\mathbf{f}$

where $\boldsymbol{\pi}_{\mathbf{if}}$ is the probability that a respondent in 'Type of activity' category i has the $\mathbf{f}^{\mathbf{th}}$ type of Further Education Desire, f $\neq \mathbf{F}$, and type F is the Further Education Desire category "No Desire".

Thus two logits are modeled for each activity type: the logit comparing Technical/Vocational Education to No Desire and the logit comparing University/Higher Education to No Desire.

Model 1: for Type of activity i

logit(response1/response3)i=1 models the probability of response category 1 relative to the response category 3.

logit(response2/response3)i=2 models the probability of response category 2 relative to the response category 3.

There are separate sets of intercept parameters $\boldsymbol{\alpha}_\mathbf{f}$ and regression parameter $\boldsymbol{\beta}_\mathbf{f}$ for each logit and the matrix xi is the explanatory variable for the i^{th} population.

Step 2: Model 1 vs Two Variable Model (Model 2)

Table 3 shows sample of data used in devising Model 2.

Table 3: Adding the 2nd Most Significant Variable to the Model

Model 2 (terms added)	*Raw Deviance (-2Log likelihood)*	*Difference in Deviance*	*Difference in df*	*p-value*
Model 1	3450.942	0	-	-
Age group	3437.718	13.2236	4	0.01023338
Gender	3434.514	16.4276	2	0.000270889
Educational level	**3288.146**	**162.7954**	**4**	**3.67573E-34**
Ethnicity	3427.465	23.4763	7	0.001407636
Province	3316.505	134.4366	16	8.79825E-21
Social class	3420.852	30.0896	6	3.77963E-05
Present financial situation	3432.545	18.3971	4	0.001031951
Financial situation in past	3448.478	2.4638	4	0.65112949
Major problems with education	3427.67	23.2713	14	0.055995893
Access to educational facilities	3437.028	13.9133	4	0.00757697

Since the variable 'Educational Level' has the lowest p-value, it was brought into the model that has been adjusted for 'Type of activity'.

<u>The explanatory variables in the model</u>: Type of activity, Educational Level

Model 2: $\mathbf{log}(\frac{\pi_{ijf}}{\pi_{ijF}}) = \boldsymbol{\alpha}_\mathbf{f} + \mathbf{x}_{\mathbf{ij}}'\boldsymbol{\beta}_\mathbf{f}$

where the matrix $\mathbf{x}_{\mathbf{ij}}$ is the set of explanatory variables for the ij^{th} population.

Step 3: Model 2 vs Three Variable Model (Model 3)

Table 4 shows sample of data used in devising Model 3.

Table 4: Adding the 3rd Most Significant Variable to the Model

Model 3 (terms added)	*Raw Deviance (-2Log likelihood)*	*Difference in Deviance*	*Difference in df*	*p-value*
Model 2	3288.146	0	-	-
Age group	3276.062	12.0842	4	0.016735994
Gender	3268.712	19.4339	2	6.02535E-05
Ethnicity	3267.805	20.3416	7	0.00487722
Province	**3155.297**	**132.8494**	**16**	**1.79279E-20**
Social class	3266.442	21.7041	6	0.001369785
Present financial situation	3276.61	11.536	4	0.02115679
Financial situation in past	3285.776	2.3706	4	0.667946717
Major problems with education	3267.495	20.6513	14	0.110907989
Access to educational facilities	3281.422	6.7239	4	0.151218299

Since the variable 'Province' has the lowest p-value, it was brought into the model already adjusted for Type of activity and Educational Level.

<u>The explanatory variables in the model</u>: Type of activity, Educational Level, Province

Model 3: $\mathbf{log}(\frac{\boldsymbol{\pi}_{\mathbf{ijkf}}}{\boldsymbol{\pi}_{\mathbf{ijkF}}}) = \boldsymbol{\alpha}_{\mathbf{f}} + \mathbf{x}'_{\mathbf{ijk}}\boldsymbol{\beta}_{\mathbf{f}}$

where the matrix $\mathbf{x_{ijk}}$ is the set of explanatory variables for the *ijk*th population.

Step 4: Model 3 vs Four Variable Model (Model 4)

Table 5 shows sample of data used in devising Model 4.

Table 5: Adding the 4th Most Significant Variable to the Model

Model 4 (terms added)	*Raw Deviance (-2Log likelihood)*	*Difference in Deviance*	*Difference in df*	*p-value*
Model 3	3155.297	0	-	-
Age group	3145.734	9.5632	4	0.048464702
Gender	**3133.988**	**21.3087**	**2**	**2.3598E-05**
Ethnicity	3152.673	2.6242	7	0.917458009
Social class	3128.48	26.8167	6	0.000156716
Present financial situation	3146.391	8.9055	4	0.063505426
Financial situation in past	3152.922	2.3747	4	0.667204138
Major problems with education	3138.381	16.9154	14	0.260714267
Access to educational facilities	3140.936	14.3609	4	0.006227982

In Step 4, Gender was found significant and brought into the model.

<u>The explanatory variables in the model</u>: Type of activity, Educational Level, Province, Gender

Model 4: $\mathbf{log}(\frac{\boldsymbol{\pi}_{\mathbf{ijklf}}}{\boldsymbol{\pi}_{\mathbf{ijklF}}}) = \boldsymbol{\alpha}_{\mathbf{f}} + \mathbf{x}'_{\mathbf{ijkl}}\boldsymbol{\beta}_{\mathbf{f}}$

where the matrix $\mathbf{x_{ijkl}}$ is the set of explanatory variables for the *ijkl*th population.

Step 5: Model 4 vs Five Variable Model (Model 5)

Table 6 shows sample of data used in devising Model 5.

Table 6: Adding the 5th Most Significant Variable to the Model

Model 5 (terms added)	*Raw Deviance (-2Log likelihood)*	*Difference in Deviance*	*Difference in df*	*p-value*
Model 4	3133.988	0	-	-
Age group	3124.457	9.5309	4	0.049116179
Ethnicity	3131.679	2.3095	7	0.940746391
Social class	**3108.071**	**25.9174**	**6**	**0.00023067**
Present financial situation	3125.265	8.7235	4	0.068394724
Financial situation in past	3131.218	2.7702	4	0.596987596
Major problems with education	3117.493	16.495	14	0.284088473
Access to educational facilities	3118.975	15.0128	4	0.004674743

In Step 5, ‘Social class’ was found significant and brought into the model.

<u>The explanatory variables in the model</u>: Type of activity, Educational Level, Province, Gender, Social class

Model 5: $\mathbf{log}(\frac{\boldsymbol{\pi}_{\mathbf{ijklmf}}}{\boldsymbol{\pi}_{\mathbf{ijklmF}}}) = \boldsymbol{\alpha}_{\mathbf{f}} + \mathbf{x}'_{\mathbf{ijklm}}\boldsymbol{\beta}_{\mathbf{f}}$

where the matrix $\mathbf{x_{ijklm}}$ is the set of explanatory variables for the *ijklm*th population.

Step 6: Model 5 vs Six Variable Model (Model 6)

Table 7 shows sample of data used in devising Model 6.

Table 7: Adding the 6th Most Significant Variable to the Model

Model 6 (terms added)	*Raw Deviance (-2Log likelihood)*	*Difference in Deviance*	*Difference in df*	*p-value*
Model 5	3133.988	0	-	-
Age group	**3124.457**	**9.5309**	**4**	**0.049116179**
Ethnicity	3131.679	2.3095	7	0.940746391
Present financial situation	3125.265	8.7235	4	0.068394724
Financial situation in past	3131.218	2.7702	4	0.596987596
Major problems with education	3117.493	16.495	14	0.284088473
Access to educational facilities	3118.975	15.0128	4	0.004674743

In Step 6, ‘Age group’ was found significant and brought into the model.

<u>The explanatory variables in the model</u>: Type of activity, Educational Level, Province, Gender, Social class, Age group

Model 6: $\mathbf{log}(\frac{\boldsymbol{\pi}_{\mathbf{ijklmnf}}}{\boldsymbol{\pi}_{\mathbf{ijklmnF}}}) = \boldsymbol{\alpha}_{\mathbf{f}} + \mathbf{x}'_{\mathbf{ijklmn}}\boldsymbol{\beta}_{\mathbf{f}}$

where the matrix $\mathbf{x_{ijklmn}}$ is the set of explanatory variables for the *ijklmn*th population.

It was observed that addition of the remaining variables did not improve the results. Hence the Model 6 has been identified as the best main effect model.

C. Improving the Model

It was further investigated to determine if the addition of two way interaction terms improved the model. The importance of an interaction term was assessed by checking the impact of the difference in deviance of the model.

Step 7: Model 6 Vs Model 7

Table 8 shows sample of data used in devising Model 7.

Table 8: Adding the 1st Most Significant 2-Way Variable to the Model

Model 7 (interaction term added)	***Raw Deviance***	***Difference in Deviance***	***Difference in df***	***p-value***
Model 6	3108.0707	0	0	-
acti*class	3077.8665	30.2042	23	0.143629043
acti*age	3073.7177	34.353	17	0.007557411
edu*pro	**2975.2098**	**132.8609**	**3**	**1.30766E-28**
edu*gen	3098.4356	9.6351	2	0.008086575
pro*class	3072.2071	35.8636	34	0.381101733
pro*age	3076.8779	31.1928	33	0.557285717
gen*class	3095.0509	13.0198	6	0.042722597
gen*age	3098.6612	9.4095	8	0.308936646
class*age	3093.4829	14.5878	16	0.555009994
eth*mprob	3055.5054	52.5653	45	0.204371948
eth*faci	3094.1994	13.8713	21	0.875063689
fin *mprob	3041.6852	66.3855	42	0.009620228
fin *faci	3079.7823	28.2884	16	0.029200335
finp*mprob	3056.0176	52.0531	43	0.162103659
finp*faci	3084.6956	23.3751	16	0.104069216
mprob *age	3051.2768	56.7939	42	0.063390185
mprob*faci	3052.4179	55.6528	44	0.111883885
faci*gen	3093.6093	14.4614	8	0.070504003
faci*class	3084.5775	23.4932	14	0.052702948
faci*age	3082.1476	25.9231	16	0.055119929

The interaction between 'Education Level' and Province has been found to have an effect on the model and hence added to the model.

<u>The explanatory variables in the model</u>: Type of activity, Educational Level, Province, Gender, Social class, Age group, Edu*Pro

Model 7: $\mathbf{log}(\frac{\boldsymbol{\pi}_{\mathbf{ijklmnf}}}{\boldsymbol{\pi}_{\mathbf{ijklmnF}}}) = \boldsymbol{\alpha}_{\mathbf{f}} + \mathbf{x}'_{\mathbf{ijklmn}}\boldsymbol{\beta}_{\mathbf{f}} + \mathbf{y}'_{(\mathbf{jk})}\boldsymbol{\beta}_{\mathbf{f}}$

where the matrix $\mathbf{x}_{\mathbf{ijklmn}}$ is the set of explanatory variables for the *ijklmn*th population and $\mathbf{y}_{(\mathbf{jk})}$ is the interaction effect of Edu*Pro.

Step 8: Model 7 Vs Model 8

Table 9 shows sample of data used in devising Model 8.

Table 9: Adding the 2nd Most Significant 2-Way Variable to the Model

Model 8 (interaction term added)	***Raw Deviance***	***Difference in Deviance***	***Difference in df***	***p-value***
Model 8	2975.2098	0	0	-
acti*age	2942.6119	32.5979	21	0.05087
edu*gen	2970.8828	4.327	2	0.114922
edu*class	2942.4092	32.8006	3	3.55E-07
pro*gen	2962.5509	12.6589	11	0.316203
pro*class	2940.9991	34.2107	27	0.160007
gen*class	2960.1934	15.0164	4	0.004667
gen*age	2961.1957	14.0141	8	0.081399
class*age	2954.729	20.4808	15	0.154254
eth*mprob	2922.7682	52.4416	42	0.129663
eth*faci	2960.932	14.2778	18	0.710811
fin *mprob	2906.8217	68.3881	40	0.003421
fin *faci	2945.8832	29.3266	14	0.009437
finp*age	2952.1219	23.0879	15	0.082291
finp*mprob	**2852.9844**	**122.2254**	**42**	**9.18E-10**
finp*faci	2952.4005	22.8093	15	0.088274
mprob *age	2913.2762	61.9336	42	0.024235
mprob*faci	2923.8426	51.3672	47	0.306598
faci*class	2955.1103	20.0995	13	0.092757
faci*age	2944.8296	30.3802	15	0.010622

In this step, the interaction between 'Financial Situation in Past (Finp)' and 'Major Problems with Education (Mprob)' has been found to have an effect on the model and hence added to the model.

<u>The explanatory variables in the model</u>: Type of activity, Educational Level, Province, Gender, Social class, Age group, Edu*Pro, Finp*Mprob

Model 8:

$$\mathbf{log}(\frac{\boldsymbol{\pi}_{\mathbf{ijklmnf}}}{\boldsymbol{\pi}_{\mathbf{ijklmnF}}}) = \boldsymbol{\alpha}_{\mathbf{f}} + \mathbf{x}'_{\mathbf{ijklmn}}\boldsymbol{\beta}_{\mathbf{f}} + \mathbf{y}'_{(\mathbf{jk})}\boldsymbol{\beta}_{\mathbf{f}} + \mathbf{z}'_{(\mathbf{op})}\boldsymbol{\beta}_{\mathbf{f}}$$

where the matrix $\mathbf{x}_{\mathbf{ijklmn}}$ is the set of explanatory variables for the *ijklmn*th population, the matrix $\mathbf{y}_{(\mathbf{jk})}$ is the interaction effect of Edu*Pro and the matrix $\mathbf{z}_{(\mathbf{op})}$ is the interaction effect of Finp*Mprob.

Step 9: Model 8 Vs Model 9

Table 10 shows sample of data used in devising Model 9.

Table 10: Adding the 3rd Most Significant 2-Way Variable to the Model

Model 9 (interaction term added)	***Raw Deviance***	***Difference in Deviance***	***Difference in df***	***p-value***
Model 9	4027.772			
acti*age	2909.589	19.7647	17	0.286441
edu*gen	2925.853	3.5005	5	0.623312
edu*class	2894.889	34.4644	15	0.002929
pro*class	2896.654	32.6994	32	0.432465
pro*age	2908.125	21.2286	31	0.905749
gen*class	2913.901	15.4524	6	0.017015
gen*age	2928.941	0.413	4	0.981399
class*age	2922.754	6.6001	12	0.882871
eth*acti	2899.069	30.2848	19	0.048284
eth*edu	2915.022	14.3318	3	0.002487
fin*acti	2907.075	22.2787	6	0.001078
fin *edu	2918.808	10.5456	11	0.482078
fin *pro	2892.44	36.9141	33	0.292752
finp*acti	2909.158	20.1959	16	0.211527
finp*edu	2919.167	10.1871	14	0.74838
mprob *edu	2892.521	36.8324	33	0.295967
mprob *pro	**2788.718**	**140.6361**	**87**	**0.00024**
mprob *gen	2895.19	34.164	34	0.459862
faci*class	2908.176	21.1779	13	0.069484
faci*age	2909.906	19.4477	11	0.053516

In this step, the interaction between 'Major Problems with Education (Mprob)' and 'Province (Pro)' has been found to have an effect on the model and hence added to the model.

<u>The explanatory variables in the model</u>: Type of activity, Educational Level, Province, Gender, Social class, Age group, Edu*Pro, Finp*Mprob, Mprob*Pro

Model 9:

$$\mathbf{log}(\frac{\boldsymbol{\pi}_{\mathbf{ijklmnf}}}{\boldsymbol{\pi}_{\mathbf{ijklmnF}}}) = \boldsymbol{\alpha}_{\mathbf{f}} + \mathbf{x}'_{\mathbf{ijklmn}}\boldsymbol{\beta}_{\mathbf{f}} + \mathbf{y}'_{(\mathbf{jk})}\boldsymbol{\beta}_{\mathbf{f}} + \mathbf{z}'_{(\mathbf{op})}\boldsymbol{\beta}_{\mathbf{f}} + \mathbf{t}'_{(\mathbf{ko})}\boldsymbol{\beta}_{\mathbf{f}}$$

where the matrix $\mathbf{x}_{\mathbf{ijklmn}}$ is the set of explanatory variables for the ijklmnth population, the matrix $\mathbf{y}_{(\mathbf{jk})}$ is the interaction effect of Edu*Pro, the matrix $\mathbf{z}_{(\mathbf{op})}$ is the interaction effect of Finp*Mprob and the matrix $\mathbf{t}_{(\mathbf{ko})}$ is the interaction effect of Mprob*Pro.

After Step 9 no more significant two-way terms were revealed.

D. Checking the Adequacy of the Best 2-Way Interaction Model

Goodness of Fit: Hypothesis Testing with Deviance Statistics

H0: No lack of fit

H1: There is some lack of fit

Table 11: Goodness of Fit Testing

	Chi-Square	**df**	**p-value**
Deviance	2145.881	2930	1.000

According to the p-value, it can be concluded that the Null Hypothesis is not rejected. Hence, the test is the proof that there is no lack of fit of the model or the model developed fits the data well.

E. Classification Table

The classification table was used to evaluate the predictive accuracy of the regression model.

Table 12: Classification Table for Model 9

OBSERVED	**PREDICTED**			
	NO DESIRE	TECHNICAL/ VOCATIONAL EDUCATION	UNIVERSITY/ HIGHER EDUCATION	% CORRECT
NO DESIRE	641	141	51	77.0%
TECHNICAL/ VOCATIONAL EDUCATION	249	304	89	47.4%
UNIVERSITY /HIGHER EDUCATION	46	45	519	85.1%
OVERALL PERCENTAGE	44.9%	23.5%	31.6%	70.2%

From the results shown in Table 12, it can be seen that Model 9 has the accuracy of more than 70 percent.

F. Implementation of Data Mining Techniques

Finally the model (Model 9) developed using Univariate analysis was used to develop a data mining model. This data mining model can predict the 'Further Educational Desire' in a youth from other attributes discussed above.

Construction of Decision Tree

A decision tree was constructed using the attributes identified as significant in the Univariate analysis. The ordering of attributes in the decision tree was also as determined in the statistical analysis. Table 13 shows the attributes and the no. of levels of each attribute. 'Type of Further Education Desire' is the class attribute.

Table 13: List of Attributes used in the Decision Tree

Seq. No	**Attribute Name**	**No. of Levels**
1	Type of Activity	7
2	Educational Level	3
3	Province	9
4	Gender	2
5	Social Class	4
6	Age group	3
7	Financial Situation in Past	3
8	Major Problems with Education	8
9	Type of Further Education Desire	3

Figure 1 shows the portion of the decision tree thus constructed.

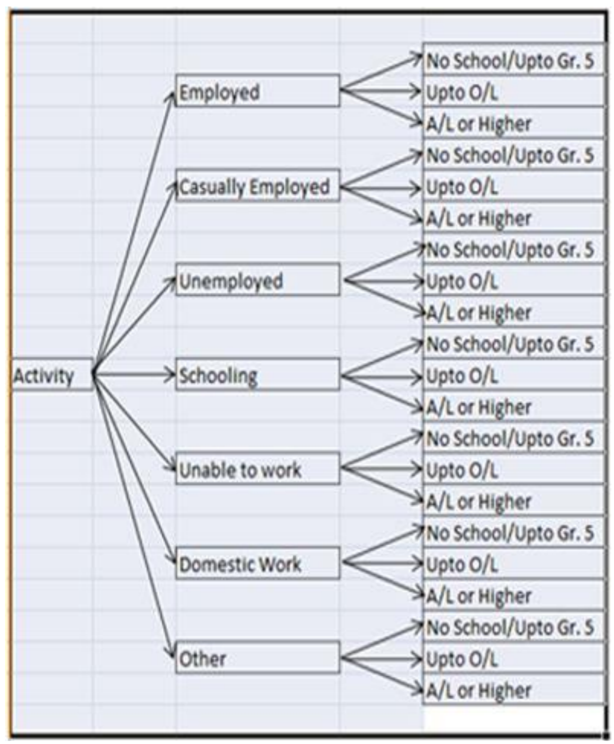


Figure 1: Portion of the Decision Tree

Figure 2 shows a sample classification rule set developed using the Decision Tree shown in Figure 1.

Rule 1: Type of Activity=Permanently Employed ^ Educational Level=No Schooling/Grade 1-5 ^ Province=Western ^ Gender=Male ^ Social Class=Middle Class ^ Age Group=20-24 yrs → No Desire
Rule 2: Type of Activity=Permanently Employed ^ Educational Level=No Schooling/Grade 1-5 ^ Province=Western ^ Gender=Male ^ Social Class=Working Class ^ Age Group=15-19 yrs → No Desire
Rule 3: Type of Activity=Permanently Employed ^ Educational Level=No Schooling/Grade 1-5 ^ Province=Central ^ Gender=Male ^ Social Class=Working Class ^ Age Group=20-24 yrs → No Desire
Rule 4: Type of Activity=Permanently Employed ^ Educational Level=No Schooling/Grade 1-5 ^ Province=Central ^ Gender=Female ^ Social Class=Working Class ^ Age Group=20-24 yrs → No Desire
Rule 5: Type of Activity=Permanently Employed ^ Educational Level=No Schooling/Grade 1-5 ^ Province=Uva ^ Gender=Male ^ Social Class=Working Class ^ Age Group=24-29 yrs → Technical/Vocational Education

Figure 2: Sample Rule Set

G. *Implementation*

A software tool was developed using Visual Basic to implement the rule set developed above. Figure 3 shows the interface of the software tool developed.

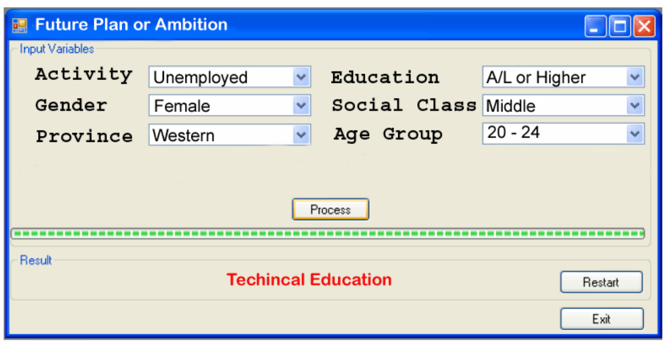


Figure 3: Interface of the Software Tool Developed

H. *Evaluation*

The system developed was tested using a random set of data containing 485 test records. The selected random set was divided into four test sets each test set containing around 125 records. Table 14 shows the classification table obtained by inputting data into the application.

Table 14: Classification Table Obtained from Test

OBSERVED	PREDICTED											
	TECHNICAL / VOCATIONAL EDUCATION				UNIVERSITY/ HIGHER EDUCATION				NO DESIRE			
TECHNICAL/ VOCATIONAL EDUCATION	21	13	19	22	3	5	5	7	13	17	16	7
UNIVERSITY /HIGHER EDUCATION	0	4	3	7	30	23	35	26	3	5	2	2
NO DESIRE	10	5	11	8	4	7	2	0	32	51	27	40

Table 15 shows the Confusion Matrices for various adjusted outcomes.

Table 15 (a): Confusion Matrix Adjusted for No Desire

			Predicted	
			No Desire	Have a Desire
Observed	Data Set 1	No Desire	36	14
		Have a Desire	16	54
	Data Set 2	No Desire	51	12
		Have a Desire	22	45
	Data Set 3	No Desire	27	13
		Have a Desire	18	62
	Data Set 4	No Desire	40	8
		Have a Desire	9	62

Table 15 (b): Confusion Matrix Adjusted for Uni/Higher Education

			Predicted	
			Uni/Higher	Not So
Observed	Data Set 1	Uni/Higher	30	3
		Not So	7	76
	Data Set 2	Uni/Higher	23	9
		Not So	12	86
	Data Set 3	Uni/Higher	35	5
		Not So	7	73
	Data Set 4	Uni/Higher	26	9
		Not So	7	77

Table 15 (c): Confusion Matrix Adjusted for Tech/Voc. Education

			Predicted	
			Tech/Voc	Not So
Observed	Data Set 1	Tech/Voc	21	16
		Not So	10	69
	Data Set 2	Tech/Voc	13	22
		Not So	9	86
	Data Set 3	Tech/Voc	19	21
		Not So	14	66
	Data Set 4	Tech/Voc	22	14
		Not So	15	68

Table 16 shows the resulting measures obtained from the tests.

Table 16: Measures Obtained from Tests

		Tech/Voc	Uni/High	No Desire	Avg.
Data Set 1	TPR	0.567568	0.909091	0. 72	0. 7322
	FPR	0.126582	0.084337	0.228571	0.1464
	Accuracy	0.775862	0.913793	0. 75	0.8132
Data Set 2	TPR	0.371429	0. 71875	0.809524	0.6332
	FPR	0.094737	0.122449	0.328358	0.1818
	Accuracy	0.761538	0.838462	0. 738462	0. 7794
Data Set 3	TPR	0.475	0.875	0.675	0.675
	FPR	0.175	0.0875	0.225	0.1625
	Accuracy	0. 708333	0.9	0. 741667	0. 7833
Data Set 4	TPR	0.611111	0. 742857	0.833333	0. 7291
	FPR	0.180723	0.083333	0.126761	0.1302
	Accuracy	0. 756303	0.865546	0.857143	0.8263
Overall	TPR	0.50628	0.81142	0. 75946	0.6923
	FPR	0.14426	0.0944	0.22717	0.1552
	Accuracy	0. 75051	0.87945	0. 77182	0.8005

From Table 16, it can be seen that the overall accuracy of the system is above 80 percent. Figure 7 shows the Receiver-Operator Characteristic (ROC) curve drawn from the data in Figure 4.

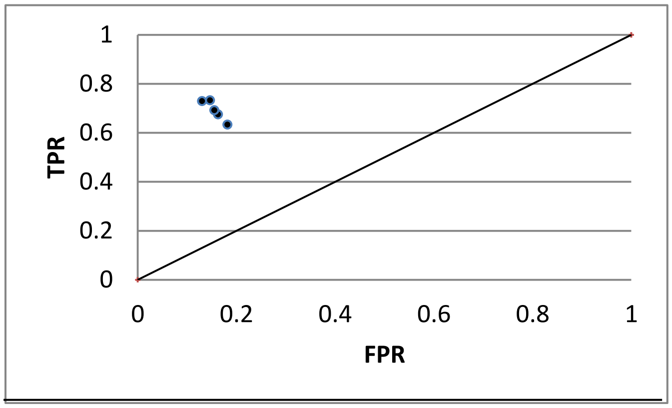


Figure 4: Receiver Operator Characteristic Curve for Model

From the ROC curve, it can be seen that the results concentrate towards the upper left hand corner. This is the proof that the accuracy of the Data Mining model is acceptable as pure random guess would lie along the diagonal line.

IV. Conclusions

This paper presented the results of the research carried out to find the factors on which the Educational Desires of Sri Lankan Youth depends. The research found that the Educational Desires of the Youth could be predicated through the combination of several social factors. The findings of the research were finally used to design a data mining model for the predication of the Educational Desires of the Youth. This model can be used by decision makers in dealing with issues concerning youth especially their further educational requirements.

References

[1] Hettige ST, Mayer M, “Sri Lankan Youth – Challenges and Responses”, Friedrich Ebert Stiftung, Colombo, Sri Lanka, (2002).

[2] Fernando HC, Tissera WMR, Athauda RI, “Gaining Insights to the Tea Industry of Sri Lanka using Data Mining”, Proceedings of the International Multi Conference of Engineers and Computer Scientists (IMECS 2008) Vol I, Hong Kong, (2008).

[3] Research Methods Knowledge Base, “Descriptive Statistics”, http://www.socialresearchmethods.net/kb/ statdesc.php [Accessed on 08/04/2010]

[4] Babbie ER, “The Practice of Social Research", 12th edition, Wadsworth Publishing, (2009).

[5] Cooper RA, Weekes AJ, “Data, Models, and Statistical Analysis”, Rowman & Littlefield Publishers, Inc., USA, (1983).

[6] Wikipedia, “Pearson's Chi-Square Test”, http://en.wikipedia.org/wiki/Pearson's_chi-square_test [Accessed on 08/04/2010]

[7] Kinley L, "Small-sample comparisons of exact levels for chi-squared goodness-of-fit statistics", Journal of the American Statistical Association 73 (362): pp. 253–263. (1978).

[8] Mehta CR, Patel NR, Tsiatis AA, "Exact Significance Testing to Establish Treatment Equivalence with Ordered Categorical Data". Biometrics 40 (3): pp. 819–825, (1984).

[9] Weisstein EW, "Fisher's Exact Test." From MathWorld--A Wolfram Web Resource. http://mathworld.wolfram.com/FishersExactTest.html [Accessed on 08/04/2010]

[10] Sarkar SK, Midi H, “Optimization Techniques for Variable Selection in Binary Logistic Regression Model Applied to Desire for Children Data” Journal of Mathematics and Statistics 5 (4): pp. 387-394, (2009).

[11] Pahwa P, Karunanayake CP, “Modeling of Longitudinal Polytomous Outcome from Complex Survey Data - Application to Investigate an Association between Mental Distress and Non-Malignant Respiratory Diseases”, BMC Medical Research Methodology, Vol 9, (2009).

Authors Profile

Mohamed Fazil Mohamed Firdhous is a senior lecturer attached to the Faculty of Information Technology of the University of Moratuwa, Sri Lanka. He received his BSc Eng., MSc and MBA degrees from the University of Moratuwa, Sri Lanka, Nanyang Technological University, Singapore and University of Colombo Sri Lanka respectively. In addition to his academic qualifications, he is a Chartered Engineer and a Corporate Member of the Institution of Engineers, Sri Lanka, the Institution of Engineering and Technology, United Kingdom and the International Association of Engineers. Mohamed Firdhous has several years of industry, academic and research experience in Sri Lanka, Singapore and the United States of America.

Ravindi Jayasundara is a lecturer attached to the Department of Mathematics, Faculty of Engineering of the University of Sri Lanka. She completed the BSc (Hons) degree in Statistics with Second Class (Upper Division) at the University of Colombo, Sri Lanka and MSc in Operations Research at the University of Moratuwa, Sri Lanka. Ravindi’s research interests are in statistics, operations research and data mining.

An Authorization Mechanism for Access Control of Resources in the Web Services Paradigm

Rajender Nath
Department of Computer Science & Applications,
Kurukshetra University, Kurukshetra,
Haryana, India

Gulshan Ahuja
Tilak Raj Chadha Institute of Management. &
Technology,
Yamunanagar,
Haryana, India

***Abstract*— With the increase in web based enterprise services, there is an increasing trend among business enterprises to migrate to web services platform. Web services paradigm poses a number of new security challenges, which can only be realized by developing effective access control models. The fact that the enterprises allow access to the resources through web services requires development of access control models that can capture relevant information about a service requester at the time of access request and incorporate this information for making effective access control decisions. Researchers have addressed many issues related to authentication and authorization of web services requests for accessing resources, but the issues related to authorization work and identity based access are still poorly addressed. Authors of this paper focus on providing an extended approach to capture relevant information about a service requester and establish a certain level of trust so that amount of authorization work required for accessing any resource is reduced and service requests are served in an efficient manner. Compared with existing access control mechanisms, the proposed mechanism has reduced the amount of authorization work required for accessing resources across varied domains.**

Keywords- Web Services; Access Control; Authentication; Authorization.

I. INTRODUCTION

Web services are loosely coupled applications which use protocols like SOAP (Simple Object Access Protocol) for message exchange, WSDL (Web Services Description Language) for interface description, and UDDI (Universal Description, Discovery, and Integration) for service discovery and communication across different security domains in a distributed environment. Currently, the most of the business enterprises try to align their business processes with the IT infrastructure support by migrating towards web service oriented architecture (WSOA). The use of identity certificates issued by trusted third parties to assign roles to users has already been addressed. Looking at the available specifications in the web service security areas like WS-Security, WS-Trust, SAML, XACML, it is understandable that developers give more emphasis on securing web services and often ignore access to the resources. In web services paradigm, a web service operation must be granted access based on the permissions and by also considering the web service invocation parameters. The aforementioned models consider an action to be a limited set of operations like read, write, execute etc. and do not consider use of invocation parameters for making access control decisions. The problem with such mechanisms is that a significant amount of authorization work is required towards verifying the credentials for each service request. Every time, a service requester sends a request to the service provider for a desired service, its identity certificates issued by the third party are verified and access control policies are evaluated before granting access to any resource. This makes the task of enforcement of access control burdensome and time consuming. To address the above problems, we propose an authorization mechanism for access control of resources in web services paradigm. The proposed mechanism considers service request invocation parameters to derive new set of parameters and uses access percentile to access any web service. Our approach presents a new means to realize the authorization in an efficient manner. The efficiency of proposed mechanism is considered in terms of reduced authorization work and time to serve any request. Our approach makes use of generalized parameters, a subset of which have to be supplied by the service requester at the time of making a request to access any resource using a web service. Another subset of parameters can be set by system administrator as per the requirements of domain in which the access is to be granted. These parameters are used to check whether access to required web service can be granted or not. The proposed approach also deals with the concept of identity tracking so that the service requests belonging to a particular service requester can be logged and the subsequent requests from the same service requester can be served with less frequency of validation checks.

This paper has been organized into IX sections. Section II discusses the state of access control for resources in web services environment. Section III discusses the related work. Section IV discusses problem formulation. Section V proposes an authorization mechanism for access control of resources in web services paradigm. Section VI presents a n example case. Section VII presents a contrast of our approach vs. existing access control approaches. Section VIII presents performance evaluation results. Section IX concludes the paper and presents future scope and improvements.

II. AN OVERVIEW OF ACCESS CONTROL FOR RESOURCES

In web services environment, service providers willing to provide services publish their web services' interfaces in public registries like UDDI. Service requesters must search and locate the web services from this publicly accessible database. Once an appropriate service has been found, the service requester can send the request to service provider for the service by encapsulating the requester's information in SOAP message using HTTP protocol. The service provider draws out the invoking parameters from SOAP messages and initiates the process to execute the appropriate methods for granting access to the required resources. Service requester and service provider may not belong to the same domain. In this case, every time service requester needs to access a resource in different domain, it has to provide its credentials for authentication and authorization. As the number of available resources and requesting entities increases, the amount of authorization work also significantly increases. The existing role based access control (RBAC) models solve this problem by providing mapping between users and roles, permissions and roles. In order to guarantee more security for resource access, researchers have developed techniques like attribute certificates (AC) based on privilege management infrastructure (PMI) [1]. PMI allows including and revoking attributes and can contain information about the privileges or roles of a user.

Every time a service requester needs to access a resource, it must provide its identity and attribute certificates, which must be validated before granting access permission. The access to resources can be made more secure by establishing certain level of trust and the amount of authorization work can be reduced by using the established trust level.

III. Related Work-Access Control Models

In 1969, the basics about access control were described formally for the first time [3] and the concepts of subjects and objects were introduced. To control the access to objects, security policies were introduced. To access any object, the security level of subject has to be more than the security level of the specified object. These types of models are termed as lattice based access control (LBAC). LBAC models are not scalable and their use is restricted in specific scenarios. To overcome these limitations, Sandhu et al. [4] in 1996 proposed the notion of Role-Based Access Control (RBAC) for enterprise applications as a powerful and generalized approach for access of enterprise resources. Later, Sanhu presented RBAC [5], which defines four reference models for RBAC as follows: RBAC0 model defines a basic RBAC system, RBAC1 model augments RBAC0 with role hierarchies, RBAC2 adds constraints to RBAC0, while RBAC3 combines RBAC1 and RBAC2. ARBAC'97 (administrative RBAC'97) [6] have three components: 1) URA97, which is concerned with userrole assignment; 2) PRA97, which focuses on permission role assignment and is a dual of URA97; 3) RRA97, which deals with role-role assignment

With the advent of web services [7, 8], the access control mechanisms became more complex. The most significant work in relation to access control in the web services was proposed by Damiani et al. [9].

Freudenthal et al. [10] presented a distributed role-based access control model (dRBAC) for dynamic coalition environments. dRBAC presents a scalable, decentralized trust-management and access control mechanism for systems that span multiple administrative domains, which utilizes PKI identities to define trust domains, roles to define controlled activities, and role delegation across domains to represent permissions to these activities.

Xu et al. [11] proposed a context aware access control model for web services. Miao Liu et al. [12] proposed an attribute and role based access control model for web services. Yizho Zhao et al. [13] proposed a model a flexible role and resource based access control model (RRBAC) which can support open and distributed environments. RRBAC is suitable for multiple security domains with different applications. RRBAC adopts role directory structure instead of role hierarchy but do not consider about authorization of requests using services.

Xu Feng et al.[14] proposed a security architecture for web services and employs a SOAP proxy for processing of SOAP messages. R Bhatti et al.[15] highlights a trust based context aware access control model for web services. Christian Emig et al. [16] proposed an access control metamodel for web service oriented architecture.

IV. Problem Formulation

A significant amount of research has been carried towards access control in web services paradigm. We have based our research on RBAC because it brings much convenience to system administrators and service requesters by introducing the concepts of role in the system. However, RBAC do not fit well for multi security domains and cannot support trust management (TM) [17] and trust negotiation (TN) [18]. Many papers have proposed frameworks to integrate RBAC with web services and addressed issues like security, context based access, attribute based access, trust establishment etc. With the increase in number of service requesters and available resources, the amount of authorization work also increases.

Security architecture [14] for web services employs a SOAP proxy for processing of SOAP messages. The architecture considers only parsing of messages and realization of a role for service access. It lacks to consider about the amount of authorization work required for web service access. The work presented by R Bhatti et al.[15] highlights a trust based context aware access control model for web services but do not consider how trust level can be established and used for access control. Christian Emig et al. [16] proposed an access control metamodel for web service oriented architecture. Their approach proposes to use an authorization verification service but do not highlight that how service requesters parameters can be utilized for reduced authorization effort.

We derive a motivation here to introduce the mechanism to use service invocation parameters like requester's credentials - identity certificates, digital signatures, object identifier etc. to establish a level of trust and conserve identity of service requests for subsequent accesses. The proposed approach uses the parameters supplied during service request and derives a new set of parameters for making access decision.

V. THE PROPOSED WORK: AN AUTHORIZATION MECHANISM FOR ACCESS CONTROL OF RESOURCES IN WEB SERVICES PARADIGM

Fig. 1, presents a model for implementing the proposed approach. The authorization approach defined by the authors of this paper uses a number of modules. The purpose and use of each module is defined as follows

A. Details about modules

1) Service Initiator Module (SIM)

The model comprises of Service Initiator Module (SIM), which is the main module for initiating the process for authorization and allowing access to the required web service.

2) Identity Tracker and Verifier (IDTV)

Identity Tracker and Verifier (IDTV) module is responsible for computation and assignment of access percentile and verification of the attribute certificates. Service Provider (SP) maintains a local database of Service Requesters for maintaining information about their identity.

3) Role Based Access Control (RBAC) processor

This module is composed of Policy Evaluator (PE) for evaluation of policies and Role Mapper (RM) for mapping service requester to assigned role. RBAC module checks service requests against specified policies from the policy store and invokes appropriate role for accessing a resource. There are three functions to perform the mapping:

- (service requester SRi, role Ri): to map a service requester SRi to a role Ri.
- (role Ri, resource RSi): role assignment, to map a role to a certain resource, i.e., relating the role Ri with resource RSi
- (resource RSi, permission set PSi): resource assignment, to map a resource to a certain permission set, i.e., corresponding the resource RSi , with the permission set PSi.

4) History Buffer (HB)

Is used for temporary conservation of information about past accesses to the resources.

B. Implementation Details for Authorization Mechanism

Service Provider (SP) maintains a local database of Service Requesters for maintaining information about their identity. To access any resource through a web service, the service requester sends a SOAP message, which includes various pieces of information about service requester to the service provider. On service provider's end, Service Initiator Module (SIM) processes it before passing it to the required web service for allowing access to the resource. Each web service is assigned an access percentile which specifies a value that must be possessed by any service request for accessing a resource through a web service.

The access percentile for each web service request is computed based on number of parameters supplied during service request. These parameters in the form of records are also maintained in history buffer. Table 2 represents a set of records for calculation of access percentile.

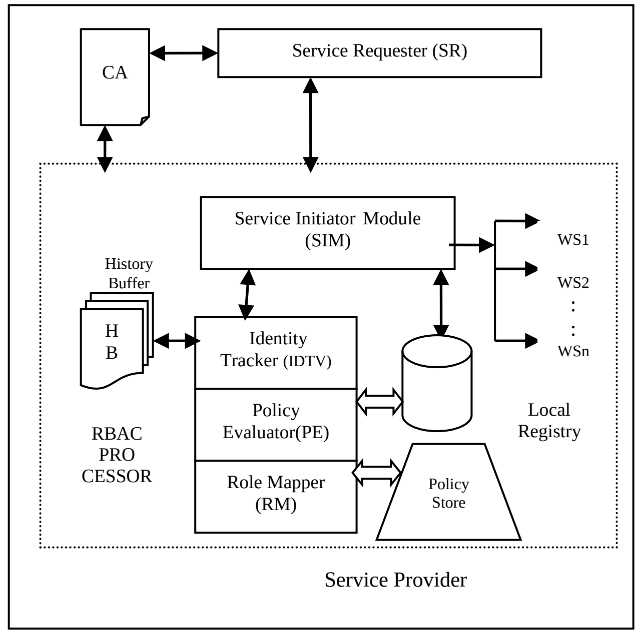


Fig. 1. Authorization architecture for access of resources using web services.

TABLE 2. RECORDS FOR ACCESS PERCENTILE

Access Parameters / Service Requesters	IP_Addr	Id _Cert	Service _ Info	Time_ to_ Live
SR_1	192.168.0.15	AXZAB-----YPZ	WS_1, WS_3	TTL_1
SR_2	192.157.12.17	QPOST-----ORP	WS_6, WS_7, WS_9	TTL_2

Columns in the table 2 represent access parameters IP_Addr, Id_cert, Service_Info, Time_to_Live fields. IP_Addr field maintains information about IP address from where the previous requests had originated. Id_Cert field maintains information about id certificate of service requester. Service_Info field maintains information about which all services have already been accessed by a service requester. Time_to_Live field states the time period for which the record in the history buffer remains alive.

The set of parameters P_rS is chosen as a generalized set, which can be changed as per requirements. Based upon these access parameters, the access percentile is computed as described below.

Each individual parameter is assigned a weight $W_{i,}$ which is assigned by the system administrator.

IP_Addr field is assigned a weight W_{IP} according to the type of IP. For e.g static IP address may be assigned more weight than dynamic IP address.

ID_Cert field is assigned weight W_{ID} as per the trust and reputation of the CA.

Service_Info field is assigned a weight $W_{IS.}$ This field keeps track of which all services have already been accessed by a service requester in a given time frame. More is the number of times a given service WS_i has been accessed the greater is the value of weight W_{IS}.

Time_to_Live field is assigned a weight W_{TTL}. The value of

TTL field decides that for how long the record will remain in the history buffer. The value for the weight W_{TTL} is computed as follows:

$W_{TTL} = RTE/TTL$

Where RTE is the remaining time within which the record will expire in the History Buffer (HB) and TTL is the total time to live in the HB. The value of weight will be more for frequently accessed services and will decrease if a service is accessed after some time gap.

Now, the access percentile for any service requester $SR_{i,Access\_Percentile}$ can be calculated as

$$SR_{i,Access\_Percentile} = X_{i,j} * W_{IP} + X_{i,j} * W_{IP} + X_{i,j} * W_{IS} + X_{i,j} * W_{TTL}.$$

$X_{i,j}$ is a presence matrix where i represents any service requester SR_i value of j ranges as $1 \leq j \leq N_{AP}$, where N_{AP} is the number of access parameters stored in HB.

$\forall SR_i \ \exists$ <IP_Addr, Id _Cert, Service_Info, Time_ to_ Live>

<IP_Addr, Id _Cert, Service_Info, Time_ to_ Live> $\asymp$ <0,0,0,0> iff SR_i does not exist in HB.

<IP_Addr, Id _Cert, Service_Info, Time_ to_ Live> $\asymp$ <1,1,1,1> iff SR_i exists in HB.

The SOAP message is checked for identity certificates of the service requester. The identity certificates can be obtained from a certified authority (CA) who is a trusted server in the domain, which generates the X.509 certificates. Each domain can have its own CA or if two or more domains trust each other then they can use a single shared CA. The primary responsibilities of CA include: Identification of clients requesting for issue of certificates, to Issue, archive and delete the certificates, to maintain a namespace of unique names for the certificate owners.

Reference Fig. 1., If SOAP message does not contain identity certificates the request is rejected otherwise local database of service requesters is checked to find out that requesting service requester already exist in the registry or not. If service requester exists in the registry, its access percentile is computed using stored parameters from HB, to check whether it can call the required web service for accessing a resource. Any service requester serving the request for the very first time will not be having any record in the history buffer. In this case, the service requester has to go through a complete cycle of validation process for granting access to any web resource through a web service.

Once a service requester has been authenticated and authorized, information about its service request parameters is recorded in the history buffer and in the local registry of registered users. The next time, if the same service requester needs to access a resource; its information is checked in the local registry, because the same service requester had already been authenticated and authorized, its record will be found in the local registry, after that the history buffer is accessed to compute access percentile based on available parameters.

If the service requester's access percentile is equal or more than the permitted access percentile, the request is passed on to RM for role mapping, otherwise the request is passed to PE for policy evaluation to cross verify against specified policies and thereafter the required resource can be accessed using the called web service. In case the computed access percentile is less than the permitted level of access percentile for required web service, the access is denied.

C. *The algorithms for the proposed mechanism are described in Fig. 2.1, Fig. 2.1(a) and Fig. 2.2*

The Service Provider (SP) will identify the decryption key and the elements to decrypt. Thereafter, each encrypted element is decrypted. If decryption fails a fault message is sent to the service requester. SIM carries the validity of SOAP message using the following method.

```
InitiateRequest( )
Input: Parameters from service requester
Output: allow,deny
{
If SR_IP € Permitted_IP
{
        if SR_i,Time_Stamp   € Permitted_Time_Stamp
        {
             if ( VerifySignature() )
                  {
                   If ( ProcessRequest() )
                       return true;
                    else
                       return false;
                  }
             else
                 return false;
        }
        else return false;
}
else return false;
}
```

Fig. 2.1 Validation of SOAP message

```
VerifySignature( )
Input: - Parameters from service requester
Output:- true,false
{
Use public key (SR_PK) of Service Requester

Digital Signature → (SR_PK) → message digest

Apply hash → message digest.

Compare and match
}
```

Fig. 2.1(a) Verification of digital signature

Verification is carried to check whether the SOAP message and enclosed credentials have come from a valid service requester or not. When service provider receives a signed SOAP message, it will initiate the verification process. The public key of the service requester is used to decrypt the digital signature and retrieve the message digest. The hash algorithm is applied again to the digital contents to generate another message digest. These two message digests are compared and if they match verification is successful. If there were any changes in the digital contents the resultant message digest would differ from the original one and the verification would fail.

```
ProcessRequest( )
Input:- Parameters from HB, Records from Local Database, Parameters from service requester
Output:-true,false
{
 If SR_i,IDCert exist in SOAP message
        Check → Local registry of SR_i where 1≤ i ≤n
        If SR_i found
        SR_i,Access_Percentile =ComputeAccessPercentile(SR_i )
                If (SR_i,Access Percentile> WS_i,Allowed_Percentile )
                        Pass SOAP message → RM
                        return true;
                end if
        elseif (ValidateCertificate(SR_i,IDCert))
                Pass SOAP message → RBAC Processor
                Update → HB
        else return false;
else return false;
}
```

Fig. 2.2 Processing of Service Requester's Information

Once the digital signatures have been verified, the SIM proceeds to check whether identity certificates are contained within message or not.

VI. An Example Case

Suppose, SecDom, be a security domain which offers online services. O_1 is the owner of a database resource system, which can provide storage service for public. O_1 registers itself for providing access to its resource through SecDom domain. After the registration, the database storage system becomes a valid resource in SecDom and the registered resource is made available through a web service. O_1 may specify many roles which are associated with different privileges. The roles include user, guest, associate_partner, admin etc. For instance, an associate partner can own the privileges of {access, read, upgrade, downgrade, delete}. Suppose a service requester wants to access a resource in SecDom domain, he authenticates himself into the domain and sends a SOAP message to the service provider, which includes service invocation parameters like requester's credentials - identity certificates and digital signatures, object identifier, security token. On service provider's end, the service request is processed for verification of supplied credentials. After that, the identity certificates are checked and validation of the certificates is carried. This is done only incase of new users or users whose request is made after a long time gap. For all those users who make request within a specific time interval, the validation of credentials can be skipped. In this way the time required to grant access to the specified resource is reduced and it also results in better bandwidth utilization.

ACL description

ACL offers a way to store the access control information. The format of ACL is chosen as follows:

acl = (role) : (resource) : (permission set)

Table 1. ACL table at service provider's end

Assigned Role	Resource Type	Permission Set
User	database storage system	{access, read, downgrade,upgrade}
Guest	database storage system	{access}
associate_partner	database storage system	{access, read, upgrade, downgrade, delete}
Admin	database storage system	{access, read, write, delete, upgrade, downgrade, create directory, delete directory, create file, delete file }

VII. Comparison With Existing Access Control Approaches

In this section, we present a comparison of our approach with Existing Access Control models based on Web Services. As per Table 3, we can find characteristics and features supported in our approach.

Table 3. Proposed Approach vs. Access Control Models for Web Services

Sr. No.	Feature/Characteristic	Web Services & Role Based Access Models	Our Approach
1.	Support for distributed environment	√	√
2.	Support for web service based access	√	√
3.	Resource request management	√	√
4.	Trust level assignment to web services		√
5.	Identity tracking & Management	√	√
6.	Reduced authorization work for access requests.		√

The above comparison reflects that our approach uses a mechanism to assign access percentile to web services and uses it to make access control decisions. The concept of identity tracking is used so that subsequent requests from same service requester can be served efficiently. The number of times the validation is to be carried for granting access to resources is reduced which results in reduced authorization work during serving of access requests.

VIII. Performance Evaluation

We consider two different scenarios for evaluation of performance of our approach.

In first scenario, we calculate time required to grant access to any resource. The measurement is carried in two ways. 1) For users who have had already requested for a resource and 2) For those users who requested for a resource for the first time. In both the cases, without our approach the time required for n service requesters would be

Following terms are defined and used for evaluation.

Let $T_{REQ} \rightarrow$ the time required to service any request made by any service requester SR_i. The time T_{REQ}, includes the round trip communication time i.e. from service requester (SR) $\rightarrow$ service provider (SP) and vice versa.

Let $T_{D,\ CERT} \rightarrow$ Time required to validate credentials of service requester from CA.

$T_{D,\ PE} \rightarrow$ Time required to evaluate associated policies.

$T_{D,\ RM} \rightarrow$ Time required to map to an appropriate role.

$T_{SRi} \rightarrow$ Total time required to serve one service request from any service requester SR_i.

Let N be the maximum number of service requests which can be made by any service requester SR_i.

$TA_{SRi} \rightarrow$ Total time required serve N service requests from same service requester SR_i

$T_{GT} \rightarrow$ Grand Total time required for granting access to all service requesters can be defined as

Let n be the maximum number of service requesters who can make a request for a resource.

We compute grand total time for first scenario as

$T_{SRi} = T_{D,\ CERT} + T_{D,\ PE} + T_{D,\ RM}$

$TA_{SRi} = N * T_{SRi}$

$T_{GT} = \sum TA_{SRi}$ where $\forall SR_i \ \ 1 \leq i \leq n$

In second scenario, we calculate time and effort required to grant access to any resource using our approach. The measurement is carried for users who have had already requested for a resource and for those users who requested for a resource for the first time.

Following terms are defined further and used for evaluation.

Let P be the number of times access percentile of any SR_i is sufficient to grant access and Q be the remaining attempts for which complete validation is required.

Also $P+Q \leq N$ and where $P \gg Q$ for all service requests made within allowed time period.

So the time required to service P service requests from same service requester SR_i within a specified time frame would be

$TP_{SRi} = P * T_{D,\ RM}$

And the time required to service Q service requests from same service requester SR_i would be

$TQ_{SRi} = Q * T_{SRi}$

Grand total time for second scenario would be

$T_{GT} = \sum(TP_{SRi} + TQ_{SRi})$ where $\forall SR_i \ \ 1 \leq i \leq n$

The above calculations reveal that once a service requester has been validated, for subsequent requests the amount of authorization work will be reduced considerably.

IX. Conclusion

In this paper, an approach has been proposed to handle authorization of service requests in web services paradigm. The approach makes use of the concept of identity tracking and access percentile for invoking any service. The paper has analyzed how the service request invocation parameters can be utilized for making efficient authorization verification. The performance of the proposed approach has been tested by taking a user case scenario. The results have indicated a good improvement in the performance during authorization of the request. The legacy systems that depend upon composition of services for accomplishing a particular task may require wrapping of these services through a common service interface, which may enable business process integration in an easy manner.

References

[1] Jordi Forne, M. Francisca Hinarejos,"Web–based Authorization based on X.509 Privilege Management Infrastructure", IEEE 2003, pp 565-568.

[2] David F. Ferraiolo, Janet A. Cugini, D. Richard Kuhn, "Role Based Access Control: Features and Motivations", Proceedings of 11th Annual Conferences, December 1995, pp 241-248.

[3] B.W. Lampson,"Dynamic protection structures",AFIPS conference proceedings, 1969, pp. 27-38.

[4] Ravi S. Sandhu, Edward J. Coyne, "Role – Based Access Control Models", IEEE Computer, Feb 1996, pp. 38-47.

[5] Ravi S. Sandhu, "Role Hierarchies and Constraints for Lattice Based Access Controls", Proceedings of the 4th European Symposium on Research in Computing Security", September 1996.

[6] R. S. Sandhu, V. Bhamidiparti, Q. Munawer. The ARBAC97 model for role-based administrator of roles. ACM Transactions on Information and System Security. Vol. 2, No. 1, Feb, 1999, pp: 105-135.

[7] Web Services Architecture, http://www.w3.org/TR/ws-arch/

[8] Web Services Activity, http://www.w3.org/2002/ws/

[9] E. Damiani, S.D.C. Vimercati, S. Paraboschi and P. Samarati, "Fine Grained Access Control for SOAP e-services", Proceedings of 10th International Conference on World Wide Web , Hong Kong, 2001, pp. 504-513.

[10] E. Freudenthal, T. Pesin, L. Port, E. Keenan, V. Karamcheti. dRBAC: disttibuted Role-Based Access Control for dynamic coalition environments. In proceeding of 21th International Conference on Distributed Computing Systems. IEEE Press, 2002, pp. 128-137.

[11] Xu F., Xie J., Huang H., and Xie L., "Context Aware Role Based Access Control Model for Web Services", Proceedings of International Workshop on Information Security and Survivability for Grid", Wuhan, China, October 21-24, 2004.

[12] Miao Liu, He Qing Guo, Jin Dian Su," An Attribute and Role Based Access Control Model for Web Services", Proceedings of the Fourt International Conference on Machine Learning Cybernetics, Guangzhou, August 18-21, 2005 pp. 1302 – 1306.

[13] Yizhu Zhao, Yanhua Zhao, Hongwei Lu,” A Flexible Role and Resource based Access Control Model”, International Colloqium on Computing, Vommunication, Control and Management, 2008, pp. 75-79.

[14] Xu Feng, Lin Guoyuan, Huang Hao, Xie Li, “Role Based Access Control System for Web Services”, Proceedings of the Fourth International Conference on Computer and Information Technology, 2004 IEEE.

[15] R. Bhatti, E. Bertino and A. Ghafoor,” A Trust Based Conext Aware Access Control Model for Web Services”, Proceedings of the IEEE International Conference on Web Services, California, USA, June6-9. 2004.

[16] Christian Emig, Frank Brandt, Sebastian Abeck,” An Access Control Metamodel for Web Service-Oriented Architecture”, International Conference on Software Engineering Advances (ICSEA 2007).

[17] X. Feng, L Jain., “Research and development of trust anagement in web security”, Journal of Software 2002, Vol. 13, No. 11, pp. 2057-2064.

[18] W. H. Winsborough, K. E. Seamons, V. E. Jones,” Automated trust negotiation. In DARPA Information Survivability Conference and Exposition”, IEEE Press, 2000, pp. 88-102.

Design and Implementation on a Sub-band based Acoustic Echo Cancellation Approach

Zhixin Chen
ILX Lightwave Corporation
Bozeman, Montana, USA
chen.zhixin.mt@gmail.com

Abstract— **This paper describes the design and implementation of a sub-band based acoustic echo cancellation approach, which incorporates the normalized least mean square algorithm and the double talk detection algorithm. According to the simulation, the proposed approach works well in the modest linear noisy environment. Since the proposed approach is implemented in fix-pointed C, it can be easily ported into fixed-point DSPs to cancel acoustic echo in real systems.**

Keywords - Digital Signal Processing; Adaptive Filter; Acoustic Echo Cancellation.

I. INTRODUCTION

People have been using telephones as a way of distant voice communication for more than one century now. In a phone conversation, echo is the sound of one's own voice being played back to him after a delay. There are two types of echo presenting in typical communication networks: hybrid echo and acoustic echo. Hybrid echo is the outcome of impedance mismatches along a telephone line, and is relatively easy to identify and cancel. Acoustic echo is due to the leakage from loudspeaker to microphone in a hands-free telephone, and is much more difficult to tackle [1].

Strong and long delayed acoustic echo signals can be very annoying, and in some cases, make conversation impossible. The effective removal of acoustic echo is thus the key to maintaining and improving voice quality on a phone call. This has led to intensive research into the area of echo cancellation, with the aim of providing solutions that can significantly reduce or even remove acoustic echo [1-3].

Good acoustic echo cancellation system can greatly enhance the audio quality of a communication system by allowing conversation to progress more smoothly and naturally, keeping the participants more comfortable, and preventing listener fatigue. However, a poorly designed acoustic echo cancellation will not provide these benefits and can even degrade audio quality significantly. This paper describes the design and simulation of an acoustic echo cancellation (AEC) approach, which is based on the sub-band method and the normalized least mean square (NLMS) algorithm, and includes the double talk detection. According to the simulation, the proposed acoustic echo cancellation approach works well in the modest linear noisy environment. Since the approach is implemented in fixed-point C language, it can be easily ported into fixed-point DSPs and used in real systems to cancel acoustic echo.

The remaining part of this paper is organized as follows. First, the principle of acoustic echo cancellation is reviewed. Next, the implementation of the proposed acoustic echo cancellation approach is described. Then, the simulation results of the proposed approach are presented. Finally, the paper is concluded with a summary of results.

II. PRINCIPLE OF ACOUSTIC ECHO CANCELLATION

The way how acoustic echo canceller works is shown in Figure 1. In general, the near end microphone signal d(n) consists of signals from near end microphone, near end background noise and echo of the far-end audio signal originating from the loudspeaker [1]. The last signal is undesired. The objective of the echo canceller is thus to form a replica of the acoustic echo signal picked up by the terminal microphone by inserting an adaptive filter h(n) parallel to the signal path through loudspeaker, room, and microphone [4]. The objective of the adaptive filter is to provide a response as equal as possible to that of the acoustic signal path. The far end speech signal x(n) is then fed through the adaptive filter h(n) to generate the signal y(n) to resemble the echo part of the microphone signal. The echo y(n) is then subtracted from the microphone signal d(n) to get the desire speech signal e(n).

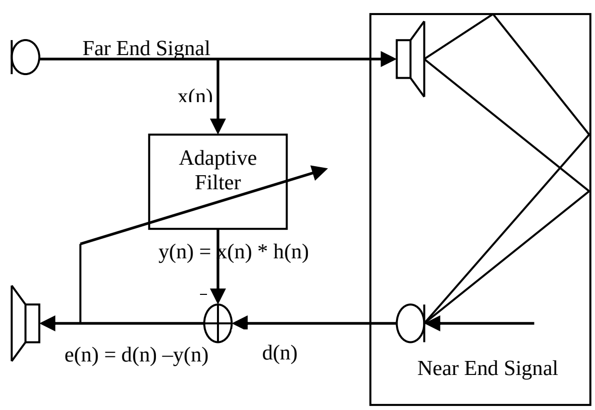


Figure 1: Principle of the acoustic echo cancellation system.

III. IMPLEMENTATION OF ACOUSTIC ECHO CANCELLATION APPROACH

From the previous description, the acoustic echo cancellation system appears as a straight-forward realization of an adaptive filtering algorithm. Some algorithms, such as LMS or NLMS, can be used in the full speech band to implement it directly [5]. However, one issue for the full-band implementation is the high computational complexity. For example, in a typical conference room, the environment echo delay varies from 100 to 500 milliseconds. For a sampling rate of 8 kHz, this means about 800 to 4000 samples and thus requires a very long adaptive filter length. Moreover, if the full band method is used, the spectrum may vary dramatically within the full band of 4 KHz [6]. Consequently, the sub-band method was used in the proposed acoustic echo cancellation approach because the spectrum will not vary a lot within each sub-band.

The LMS algorithm is the most popular algorithm in adaptive signal processing due to its simplicity and reliability. Its main drawback is that the speed of convergence gets very slow if there is a big spread among the eigenvalues [6]. Thus, the normalized NLMS algorithm is used in the proposed approach because it is insensitive to the amplitude variations of the inputs signal so that it can achieve a robust performance.

Double talk is another important issue related to acoustic echo cancellation. If the state of double talk cannot be detected accuarately, it can cause divergence of the adaptive algorithm. Consequently, a robust double-talk and single-talk detection algorithm was designed in the proposed AEC approach.

In the following sections, the sub-band approach, the LMS and NLMS algorithm, and the double talk detection algorithm will be presented.

A. Sub-band Approach

As mentioned previously, the sub-band implementation has two advantages. First, the complexity is reduced by dividing the signal into sub-bands and applying adaptive filters to a decimated signal in each sub-band. Second, the spectral variability within a sub-band is reduced compared with the full band signal [1]. To maintain the high quality of speech, the analysis and synthesis filterbanks in the sub-band realization must be designed to provide perfect reconstruction, which means that the signal fed through the analysis and synthesis filterbank system shall be an exact but delayed copy of the input signal. In the proposed AEC approach, the Quadrature Mirror Filters (QMF) was used to decompose the full band into four sub-bands with an identical bandwidth of 1 kHz. The reason that the full band is divided into four sub-bands is that the energy for speech focuses on the lowest 1 kHz range and the spectrum will be almost constant within the lowest 1 kHz sub-band. The basic architecture for a two-channel QMF system is shown in Figure 2.

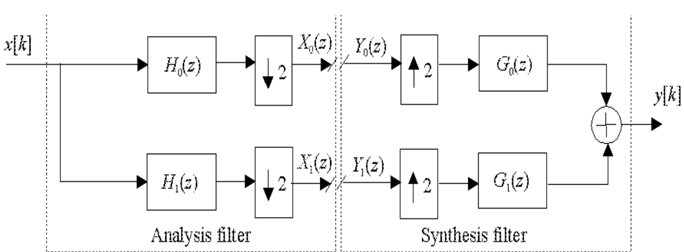


Figure 2: A block diagram of the QMF filter architecture

In Figure 2, the down arrows represent decimation by 2 and the up arrows denote interpolation by 2. The two-channel QMF system consists of two input-output paths, each of which has a bandwidth requirement that is half the original bandwidth requirements. In Figure 2, the top path contains a low-pass filter while the bottom path has a high-pass filter. Since the FIR filter is used here, a very steep filter cannot be achieved without a very large filter tap. The frequency response of the top path will have some overlap with that of the bottom. The overlap that exists beyond one-fourth the sampling frequency in the top channel, when re-sampled at half the sampling frequency, introduces aliasing errors. The bottom path also introduces aliasing errors. According to [1], the aliasing errors in the top and bottom path are self-cancelling if the following condition is met.

$$(H_0(-z)G_0(z)+H_1(-z)G_1(z))X(-z)=0\,, \tag{1}$$

which is satisfied if $G_0(z)=H_1(-z)$ and $G_1(z)=H_0(-z)$. A special case assumes that $H_0(z)$ and $H_1(z)$ are sub-band filters satisfying the mirror filter relationship $H_1(z)=H_0(-z)$ [1, 6].

Based on the above-mentioned theory, a 45th order linear phase filter that is shown in Figure 3 was used in this paper. It is shown that the low-pass and high-pass filter are mirror symmetry. This QMF was used twice to get four sub-bands.

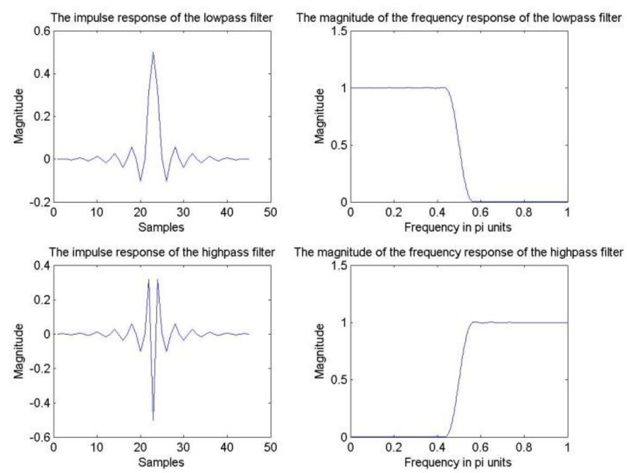


Figure 3: The impulse and frequency response of the QMF used in the proposed AEC approach

B. NLMS Algorithm

It was described previously that the goal of the AEC system is to identify an unknown system, the loudspeaker-to-microphone transfer function, with an adaptive filter. That is, the input signal needs to be processed so that it matches the desired signal $d(n)$. To measure how well the cancellation is done, a performance criterion $J(e)$ to the error is used here. According to [4], the performance criterion is chosen to be a squared error measure and it can be defined as follows.

$$J(e)=\frac{1}{2}E(e^2(n))=\frac{1}{2}E[(d(n)-y(n))^2]\,. \tag{2}$$

Here the room transfer function is assumed to be linear and wide-sense stationary and the desired and input signals are assumed to be jointly wide sense stationary and zero mean. The filter is composed by a certain number of parameters $h_n(k)$ that are set to minimize the error $J(e)$. For FIR filters, the output of the filter can be written in the form $y(n)=\sum_{k=0}^{N-1}h_n(k)x(n-k)$, (3)

where N is the filter order and $x(n)$ is the input signal. The optimal $h_n(k)$ parameters are found by solving

$$\nabla_h(J)=\nabla_h(E(e^2(n)))=0, \quad (4)$$

which leads to the Yule-Walker equation [4, 5, 7].

There are many algorithms for this equation, among which the most popular ones are LMS and NLMS. In the LMS algorithm, the adaptation rules are shown in the following equation:

$$\hat{h}_{n+1}(k)=\hat{h}_n(k)+\mu e(n)x(n-k), \quad (5)$$

where $k=1,2,\ldots,N$ and $e(n)=d(n)-y(n)$. Here $\hat{h}_{n+1}(k)$ are the filter parameters at time $n+1$ and μ is a constant adaptation step. The algorithm is proven to converge if the error-performance surface is quadratic and if the adaptation step satisfies $\mu<\frac{1}{\lambda_{\max}}$, where $\lambda_{\max}$ is the biggest eigenvalue of input correlation matrix R. Usually, the following condition is taken

$$\mu<\frac{2}{input\ power}. \quad (6)$$

The main drawback of the LMS algorithm is the speed of convergence gets very slow if there is a big spread among the eigenvalues of R. Different from LMS, the NLMS algorithm is insensitive to the amplitude variations of the input signal. The NLMS algorithm is given by

$$h_{n+1}=h_n+\beta\frac{x^*(n)}{||x(n)||^2}e(n). \quad (7)$$

According to [5], the NLMS algorithm converges in the mean-square if $0<\beta<2$. In the LMS algorithm, the correction that is applied to $h_n(k+1)$ is proportional to the input vector $x(n)$. Therefore, when $x(n)$ is large, the LMS algorithm experiences a problem with gradient noise amplification. With the normalization of the LMS step size by $||x(n)||^2$ in the NLMS algorithm, however, this noise amplification problem is diminished. However, a similar problem occurs when $x(n)$ becomes too small. An alternative way is therefore to use the following modification to the NLMS algorithm:

$$h_{n+1}=h_n+\beta\frac{x^*(n)}{\varepsilon+||x(n)||^2}e(n), \quad (8)$$

where ε is some small positive number [4, 5].

In the proposed AEC approach, the echo delay that this system can handle is set to 360 ms, which means that the taps of the adaptive FIR filters will be up to 720. Since the speech energy is different for different sub-bands, the taps of the filter are different for different sub-bands. The taps are 720, 540, 360, and 240 for the sub-bands from low to high respectively. The step size was also chosen to be larger for higher sub-bands, which mean that fast convergence is expected when the echo signal has shorter delay and has higher frequency components. This is because the speech energy focuses on the lower frequency and so does the echo.

For the updating of the filter coefficient, Eq. (8) was used. Here ε was chosen to be 10^{-8} and the step size was chosen from 10^{-5} to 10^{-4}, which is much larger than the traditional value of 10^{-7} but still gets a good result.

C. Double Talk Detection

If the coefficients of the adaptive filters are updated during double talk, the filter may be divergent. A double talk detection algorithm was thus included in the proposed AEC approach to detect the double talk and is described as follows. For both far end and near end signal, a long-term energy and a short-term energy were used to judge whether it is silent or not. The coefficients of the adaptive filters will only be updated when the far end talks and the near end does not talk because the transfer function of the room acoustics is expected in this case. Three convergence states, low, medium, and deep convergence states are defined here and shown in Figure 4. Initially, the system is in low convergence state. When neither far end nor near end talks, the system will keep the convergence state. When the far end talks and the power of the echo reduced signal is less than that of the near end signal for at least two seconds, the system will transit into medium convergence state. The system will transit into deep convergence state if this condition continues for two more seconds. If double talk stays for one second, the system will transit into the low convergence state no matter what state it is in.

If the system is in deep convergence state, the filter coefficients will not be updated since in this case the room transfer function will not change. If the system is in low or medium convergence state, the coefficients will be updated when the far end talks. The step size for low convergence state is larger than that for medium convergence state because the system is expected to converge faster initially but converge slower when it is already in the modest convergence.

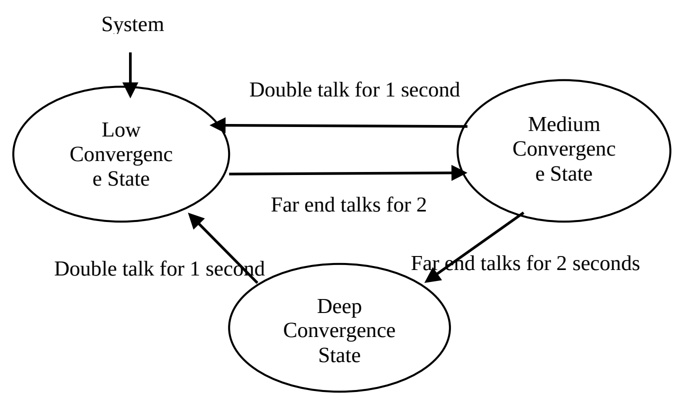


Figure 4: Transition among the three convergence states

With the single talk and double talk detection and the flexible transition of the convergence states presented above, the system performs very well in the modest noise environment.

IV. SIMULATION OF THE ACOUSTIC ECHO CANCELLATION APPROACH

The proposed acoustic echo cancellation approach was implemented using fix-point C language. The simulation results for several different scenarios are presented in the following subsections.

A. *Only Near End Talks*

In the first scenario, the far end is set to silent. This simulates the case that the microphone is turned off in the far end. The simulation is to test that the quality of the near end signal does not degrade. The system performs as expected in this case, which can be seen in Figure 5. In the figure, R_{in} means the far end signal, S_{in} denotes the near end signal, and S_{out} represents the desired near end output signal.

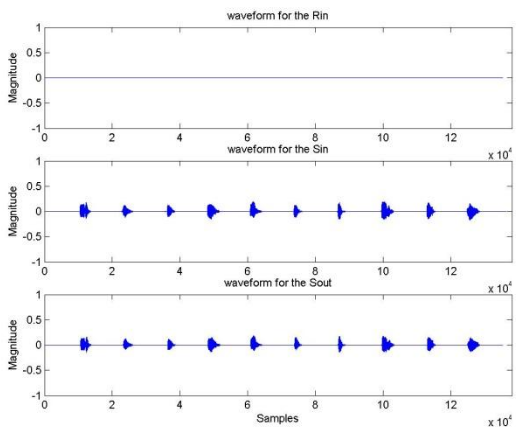


Figure 5 : Far end does not talk

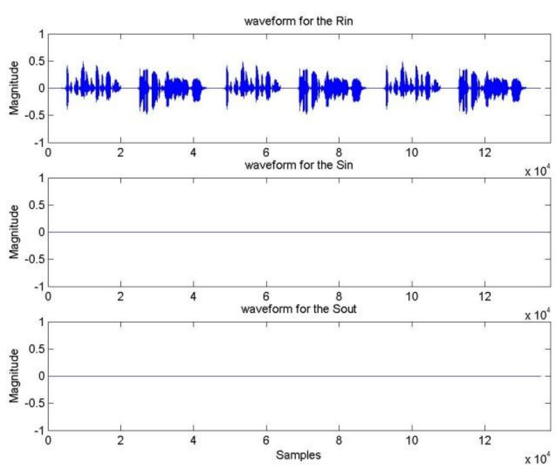


Figure 6: Microphone is turned off in the near end

B. *Only Far End Talks*

In the second scenario, the near end was set to silent, which simulates the case that the microphone is turned off in the near end. The goal of this simulation is to verify that there is no output signal. Figure 6 shows that the system performs as expected.

In the third scenario, the microphone in the near end is set to open. This means that an echo might present in the near end. The goal of this simulation is to test how well the system deletes the echo. The simulated echo signal is a 20% attenuation and 300 ms delay version of the original signal. As shown in Figure 7, the system completely removes the acoustic echo.

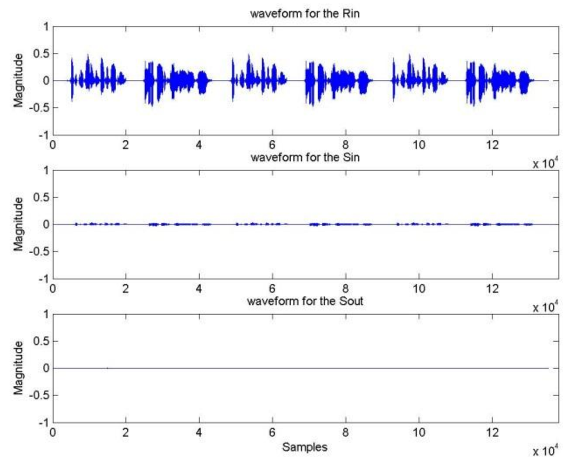


Figure 7: Near end does not talk but the microphone is open

In the following scenario, the effect of the number the sub-bands was tested. One to four sub-bands were used for the signal and it is shown that the simulation result is almost the same. It takes about 20000 samples (2.5 seconds) for the system to converge and remove the acoustic echo. The similar results for different number of used subbands are because the energy of the speech focuses on the low frequency range and so does that of the echo. Consequently, it does not matter whether the higher sub-bands were processed or not. More details can be seen in Figure 8.

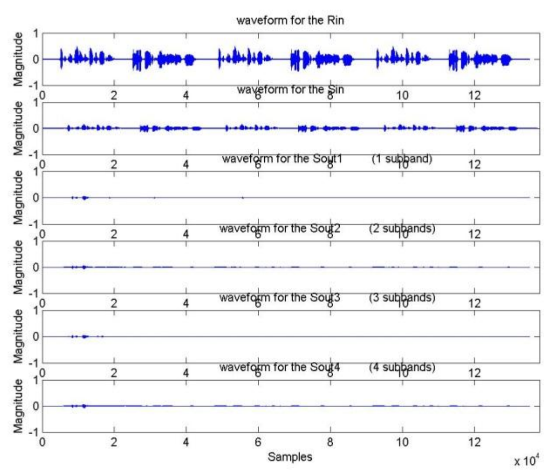


Figure 8: Effect of the number of sub-bands on the AEC system performance

In the former scenarios, the delay is constant for the entire signal. In this scenario, the delay is not constant, which means that the delay for the first sentence can be 150 ms while that for the second sentence is 200 ms. The system also works well in this case, which is shown Figure 9.

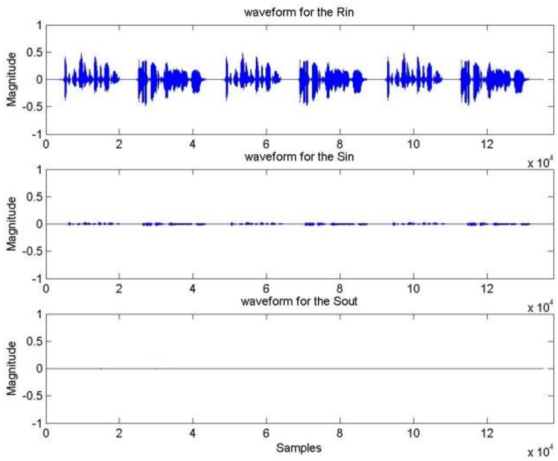


Figure 9: Simulation of irregular delay

Till now, only one echo component is considered in the simulations. In this scenario, the case that the echo includes many copies of the original signal with different delays and different gain is investigated. This means that the speech from the speaker can be reflected by the wall and then fed into the microphone in 10ms, 150ms or 200ms. It is shown in Figure 10 that the system also works well in this case.

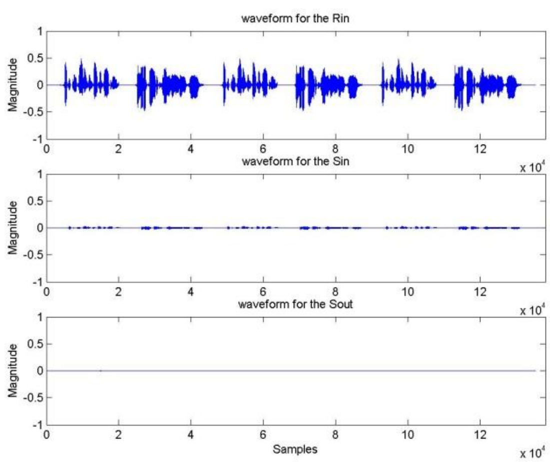


Figure 10: Simulation of many delays

C. Double Talk

In the following scenarios, both the far end and the near end are assumed to be active. A simulation was run to investigate the case that the far end and the near end talk in different time, which happens in most of the telephone communication systems. It is expected that if the far end and the near end do not talk at the same time, the performance should be as good as that of the single talk. The simulation result shown in Figure 11 verifies this.

In another simulation, some white noise was added in the near end. Since the noise is a part of the near end signal, it should not affect the performance. Figure 12 verifies the simulation result.

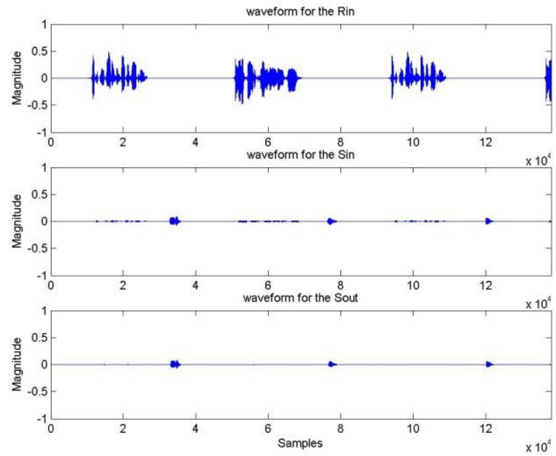


Figure 11: Far end and near end talk in different time

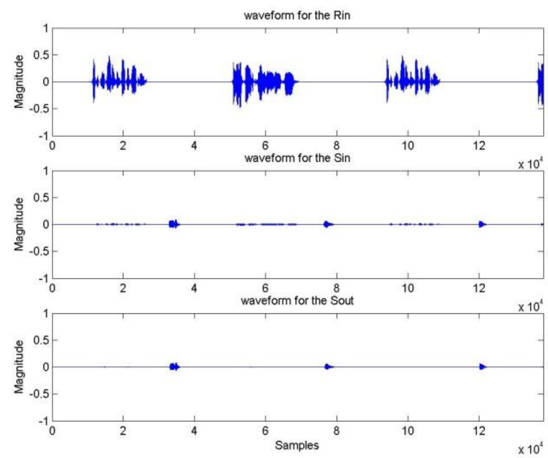


Figure 12: Noise shows up in the near end

It is assumed that the far end and the near end talk at the same time in the last scenario. Since the room transfer function changes from time to time, it is expected that the performance will degrade. Figure 13 shows the detailed simulation result.

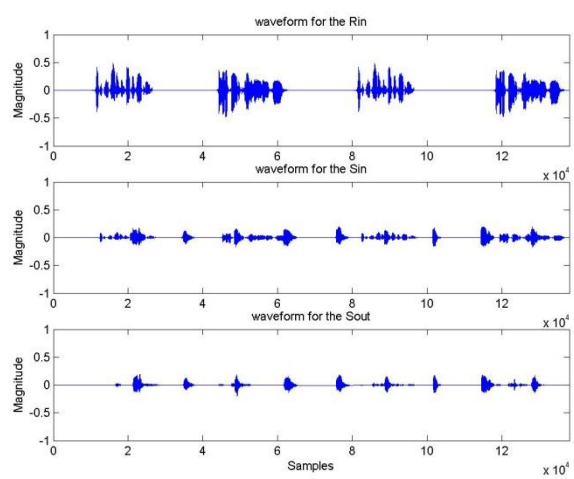


Figure 13: Far end and near end talk in the same time

V. CONCLUSION

In this paper, an acoustic echo cancellation system based on the sub-band approach and the NLMS algorithm is proposed. A double-talk detection algorithm is also included in the system. From the simulations, the system works well for the echo with 300 ms delay in modest noise environment. It works very well when only one end talks or when both end talk in the different time. But it does not work very well when both end talk in the same time. One thing deserving the attention is that this system is designed for the linear echo delay, so it is not appropriate for the environment that leads to echoes with new frequency components. An acoustic echo system which can eliminate the nonlinear echoes will be designed in the future.

REFERENCES

[1] A. Gilloire and M. Vetterli, “Adaptive filtering in subbands with critical sampling: analysis, experiments, and application to acoustic echo cancellation”, IEEE Trans. Signal Process., vol. 40, p.1862-1875, 1992

[2] J. Benesty, T. Gänsler, D. R. Morgan, M. M. Sondhi, and S. L. Gay, “Advances in network and acoustic echo cancellation”, 2001, Springer.

[3] C. Antweiler, J. Grunwald and H. Quack, “Approximation for optimal stepsize control for acoustic echo cancellation,” Proc. ICASSP-97, IEEE Conference on Acoustics Speech and Signal Processing, München, Germany, vol. 1, pp. 295-298, 1997

[4] Paolo Prandoni, “Acoustic echo cancellation” http://lcavwww.epfl.ch/~prandoni/dsp/echo/echo.html

[5] K. A. Lee and W. S. Gan "Improving convergence of the NLMS algorithm using constrained subband updates", IEEE Signal Process. Letter, vol. 11, p.736-739, 2004.

[6] G. Zelniker and F. Taylor, “Advanced digital signal processing”, Marcel Dekker, 1994

[7] M. G. Siqueira, "Adaptive filtering algorithms in acoustic echo cancellation and feedback reduction", Ph.D. thesis, University of California, Los Angeles, 1998

AUTHORS PROFILE

Zhixin Chen is a Firmware Engineer in ILX Lightwave Corporation. He holds a BS degree from Xiamen University in China, a MS degree from Xiamen University in China, and a Ph.D. degree from Montana State University in USA, all in Electrical and Computer Engineering. His research interest and working experience are in the area of acoustics, audio, and speech processing, multimedia communication, and embedded system design for high power current source, temperature controller, and optical power meter.

An Efficient Density based Improved K- Medoids Clustering algorithm

Raghuvira Pratap A
M.Tech,CSE Department
P.V.P Siddhartha Institute of Technology
Vijayawada-07, AP

K Suvarna Vani
Assistant Professor,CSE Department
V.R. Siddhartha Engineering College
Vijayawada -07, AP

J Rama Devi
Sr.Assistant Professor,CSE Department
P.V.P Siddhartha Institute of Technology
Vijayawada-07, AP

Dr.K Nageswara Rao
Professor & Head,CSE Department
P.V.P Siddhartha Institute of Technology
Vijayawada-07, AP

***Abstract*— Clustering is the process of classifying objects into different groups by partitioning sets of data into a series of subsets called clusters. Clustering has taken its roots from algorithms like k-medoids and k-medoids. However conventional k-medoids clustering algorithm suffers from many limitations. Firstly, it needs to have prior knowledge about the number of cluster parameter k. Secondly, it also initially needs to make random selection of k representative objects and if these initial k medoids are not selected properly then natural cluster may not be obtained. Thirdly, it is also sensitive to the order of input dataset.**

Mining knowledge from large amounts of spatial data is known as spatial data mining. It becomes a highly demanding field because huge amounts of spatial data have been collected in various applications ranging from geo-spatial data to bio-medical knowledge. The database can be clustered in many ways depending on the clustering algorithm employed, parameter settings used, and other factors. Multiple clustering can be combined so that the final partitioning of data provides better clustering. In this paper, an efficient density based k-medoids clustering algorithm has been proposed to overcome the drawbacks of DBSCAN and kmedoids clustering algorithms. The result will be an improved version of kmedoids clustering algorithm. This algorithm will perform better than DBSCAN while handling clusters of circularly distributed data points and slightly overlapped clusters.

Keywords- Clustering; DBSCAN; Centroid; Medoid; k-medoids.

I. Introduction

Numerous applications require the management of spatial data, i.e. data related to space. Spatial Database Systems (SDBS) (Gueting 1994) are database systems for the management of spatial data. Increasingly large amounts of data are obtained from satellite images, X-ray crystallography or other automatic equipment. Therefore, automated knowledge discovery becomes more and more important in spatial databases.

Clustering algorithms are attractive for the task of class identification. However, the application to large spatial databases raises the following requirements for clustering algorithms:

(1) Minimal requirements of domain knowledge to determine the input parameters, because appropriate values are often not known in advance when dealing with large databases.

(2) Discovery of clusters with arbitrary shape, because the shape of clusters in spatial databases may be spherical, drawn-out, linear, elongated etc.

(3) Good efficiency on large databases, i.e. on databases of significantly more than just a few thousand objects.

Clustering is considered as one of the important techniques in data mining and is an active research topic for the researchers. The objective of clustering is to partition a set of objects into clusters such that objects within a group are more similar to one another than patterns in different clusters. So far, numerous useful clustering algorithms have been developed for large databases, such as K-MEDOIDS [1], CLARANS [2], BIRCH [3], CURE [4], DBSCAN [5], OPTICS [6], STING [7] and CLIQUE [8]. These algorithms can be divided into several categories. Three prominent categories are partitioning, hierarchical and density-based. All these algorithms try to challenge the clustering problems treating huge amount of data in large databases. However, none of them are the most effective. In density-based clustering algorithms, which are designed to discover clusters of arbitrary shape in databases with noise, a cluster is defined as a high-density region partitioned by low-density regions in data space. DBSCAN (Density Based Spatial Clustering of Applications with Noise) [5] is a typical Density-based clustering algorithm. In this paper, we present a new algorithm which overcomes the drawbacks of DBSCAN and k-medoids clustering algorithms.

II. Literature Survey

A. *DBSCAN: Density Based Spatial Clustering of Applications with Noise.*

In this section, we present the algorithm DBSCAN (Density Based Spatial Clustering of Applications with Noise) which is designed to discover the clusters and the noise in a spatial database. Ideally, we would have to know the appropriate parameters Eps and MinPts of each cluster and at least one point from the respective cluster. Then, we could retrieve all points that are density-reachable from the given point using the correct parameters. But there is no easy way to get this information in advance for all clusters of the database. However, there is a simple and effective heuristic to determine the parameters Eps and MinPts of the "thinnest", i.e. least dense, cluster in the database. Therefore, DBSCAN uses global values for Eps and MinPts, i.e. the same values for all clusters. The density parameters of the “thinnest” cluster are good candidates for these global parameter values specifying the lowest density which is not considered to be noise.

The idea of it was:

1. ε-neighbor: the neighbors in ε semi diameter of an object
2. Kernel object: certain number (*MinP*) of neighbors in ε semi diameter.
3. To a object set *D*, if object *p* is the ε-neighbor of *q*, and *q* is kernel object, then *p* can get “direct density reachable” from *q*.
4. To a ε, *p* can get “direct density reachable” from *q*; *D* contains *Minp* objects; if a series object $p_1, p_2, \ldots p_n, p_1 = q.\ p_n = q.$ *then* p_{i+1} can get “direct density reachable” from p_i, $p_i \in D, 1 \leq i \leq n$.
5. To ε and *MinP*, if there exist an object $o(o \in D)$ *p* and *q* can get “direct density reachable” from *o*,*p* and *q* are density connected.

Density Reachability and Density Connectivity:

Density reachability is the first building block in dbscan. It defines whether two distance close points belong to the same cluster. Points p1 is density reachable from p2 if two conditions are satisfied: (i) the points are close enough to each other: distance (p1, p2) <e, (ii) there are enough of points in is neighborhood: |{r: distance(r, p2)}|>m, where r is a database point.

Density connectivity is the last building step of dbscan. Points p0 and pn are density connected, if there is a sequence of density reachable points p1,i2,...,i(n-1) from p_0 to p_n such that p(i+1) is density reachable from pi. A dbscan cluster is a set of all density connected points.

Explanation of DBSCAN Steps

- DBScan requires two parameters: epsilon (eps) and minimum points (minPts). It starts with an arbitrary starting point that has not been visited. It then finds all the neighbor points within distance eps of the starting point.
- If the number of neighbors is greater than or equal to minPts, a cluster is formed. The starting point and its neighbors are added to this cluster and the starting point is marked as visited. The algorithm then repeats the evaluation process for all the neighbors’ recursively.
- If the number of neighbors is less than minPts, the point is marked as noise.
- If a cluster is fully expanded (all points within reach are visited) then the algorithm proceeds to iterate through the remaining unvisited points in the dataset.

Advantages of DBSCAN

- DBScan requires two parameters: epsilon (eps) and minimum points (minPts). It starts with an arbitrary starting point that has not been visited. It then finds all the neighbor points within distance eps of the starting point.
- If the number of neighbors is greater than or equal to minPts, a cluster is formed. The starting point and its neighbors are added to this cluster and the starting point is marked as visited. The algorithm then repeats the evaluation process for all the neighbors’ recursively.
- If the number of neighbors is less than minPts, the point is marked as noise.
- If a cluster is fully expanded (all points within reach are visited) then the algorithm proceeds to iterate through the remaining unvisited points in the dataset.

Disadvantages of DBSCAN

- DBScan requires two parameters: epsilon (eps) and minimum points (minPts). It starts with an arbitrary starting point that has not been visited. It then finds all the neighbor points within distance eps of the starting point.
- DBSCAN cannot cluster data sets well with large differences in densities, since the minPts-epsilon combination cannot be chosen appropriately for all clusters then

B. K-Medoids

K-medoid is a classical partitioning technique of clustering that clusters the data set of *n* objects into *k* number of clusters [1, 3]. This k: the number of clusters required is to be given by user. This algorithm operates on the principle of minimizing the sum of dissimilarities between each object and its corresponding reference point. The algorithm randomly chooses the k objects in dataset D as initial representative objects called medoids. A medoid can be defined as the object of a cluster, whose average dissimilarity to all the objects in the cluster is minimal i.e. it is a most centrally located point in the given data set. Then for all objects in the dataset, it assigns each object to the nearest cluster depending upon the object’s distance to the cluster medoid. After every assignment of a data object to particular cluster the new medoid is decided.

1) Input

k: the number of clusters.

D: a data set containing n objects.

2) Output

A set of k clusters.

3) Algorithm

1. Randomly choose k objects in D as the initial representative objects;

2. for all objects in the data set D

a. Find the cluster C which is nearest to object i by using the dissimilarity measure;
b. assign object i to cluster C;
c. set the member object in cluster C having minimum intra cluster variance as new centroid of C

3. Display statistics of clusters obtained.

The problem is K-Medoids does not generate the same result with each run, because the resulting clusters depend on the initial random assignments. It is more robust than kmedoids in the presence of noise and outliers; however it's processing is more costly than the k-medoid method. Lastly, the optimal number of clusters k is hard to be predicted, so it is difficult for a user without prior knowledge to specify the value of k.

Problems with kmedoids clustering algorithm

The algorithm is simple and has nice convergence but there are number of problems with this. Some of the weaknesses of k-mediods are

- When the numbers of data are not so many, initial grouping will determine the cluster significantly.
- The result is circular cluster shape because based on distance.
- The number of cluster, K, must be determined beforehand. Selection of value of K is itself an issue and sometimes it's hard to predict beforehand the number of clusters that would be there in data.
- We never know the real cluster, using the same data, if it is inputted in a different order may produce different cluster if the number of data is few.
- Experiments have shown that outliers can be a problem and can force algorithm to identify false clusters.
- We never know which attribute contributes more to the grouping process since we assume that each attribute has the same weight.

III. PROPOSED ALGORITHM

Let D is the Dataset with k points
k be the number of clusters to be found
l be the number of clusters initially found by density based clustering algorithm
ε be the Euclidean neighborhood radius
ή Minimum number of neighbors required in ε neighborhood to form a cluster
p can be any point in D
N is a set of points in ε neighborhood of *p*
c=0
for each unvisited point ***p*** in dataset **D**

{
N = getNeighbors (***p***, **ε**)
if (sizeof(**N**) < ή)
mark ***p*** as NOISE
else
++ ***c***
mark ***p*** as visited
add ***p*** to cluster ***c***
recurse (**N**)
}
Now will have **m** clusters
for each detected clusters {
find the cluster centers Cmby taking the representative object.
find the total number of points in each cluster
}
If **m**>**k {**
Join two or more as follows
select two cluster based on density and number of points satisfying the application criteria and joint them and find the new cluster center and repeat it until achieving k clusters.
Finally we will have Ck centers
} else {
l =k-m
# split one or more as follows
if (**m** >=**l**) {
Select a cluster based on density and number of points satisfying the application criteria and split it using kmedoids clustering algorithm and repeat it until achieving k clusters.

Finally we will have Ck centers
}

Apply one iteration of k-medoid clustering with k and new Ck centers as the initial parameters and label all the clusters with k labels.

Note: in our simulation of the algorithm, we only assumed overlapped clusters of circular or spheroid in nature. So the criteria for splitting or joining a cluster can be decided based on the number of expected points in a cluster or the expected density of the cluster (derived by using the number of points in a cluster and the area of the cluster)

IV. EVALUATION AND RESULTS

Metrics Used For Evaluation

In order to measure the performance of a clustering and classification system, a suitable metric will be needed. For evaluating the algorithms under consideration, we used Rand Index and Run Time as two measures

A. Performance in terms of time

We evaluated the three algorithms DBSCAN, k-medoid and DBkmedoids in terms of time required for clustering. The Attributes of Multidimensional Data:

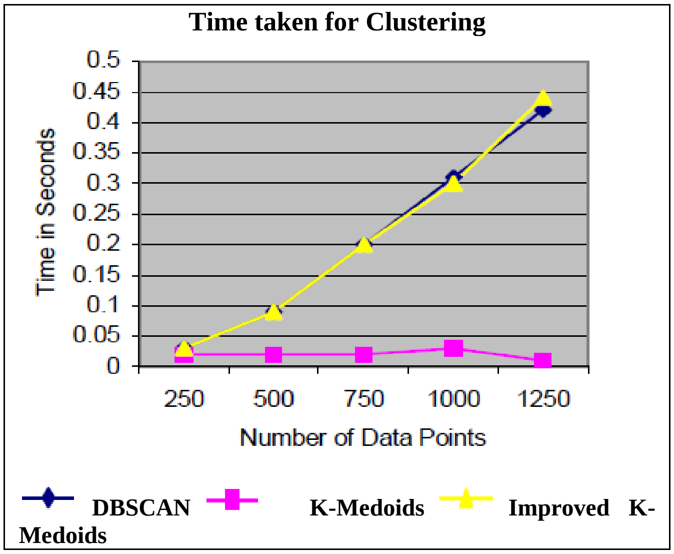
The Number of Classes: 5

The Number of Dimensions: 2

The Number of Points per Class: 50, 100, 150, 200,250

The Standard Deviation: 7.000000e-001

S.No.	Total No. of Records in Proteins Data Base	Time taken for Classification (in Seconds)		
		DBSCAN	K-Medoids	Density based Improved K-Medoids
1	250	0.03	0.02	0.03
2	500	0.09	0.02	0.09
3	750	0.2	0.02	0.2
4	1000	0.31	0.03	0.3
5	1250	0.42	0.01	0.44



B. Performance in terms of accuracy

The Rand index or Rand measure is a commonly used technique for measure of such similarity between two data clusters. This measure was found by W. M. Rand and explained in his paper "Objective criteria for the evaluation of clustering methods" in Journal of the American Statistical Association (1971).

Given a set of n objects S = {O1, ..., On} and two data clusters of S which we want to compare: X = {x1, ..., xR} and Y = {y1, ..., yS} where the different partitions of X and Y are disjoint and their union is equal to S; we can compute the following values:

a is the number of elements in S that are in the same partition in X and in the same partition in Y,

b is the number of elements in S that are not in the same partition in X and not in the same partition in Y,

c is the number of elements in S that are in the same partition in X and not in the same partition in Y,

d is the number of elements in S that are not in the same partition in X but are in the same partition in Y.

Intuitively, one can think of a + b as the number of agreements between X and Y and c + d the number of disagreements between X and Y. The Rand index, R, then becomes, The Rand index has a value between 0 and 1 with 0 indicating that the two data clusters do not agree on any pair of points and 1 indicating that the data clusters are exactly the same.

The Attributes of Multidimensional Data:

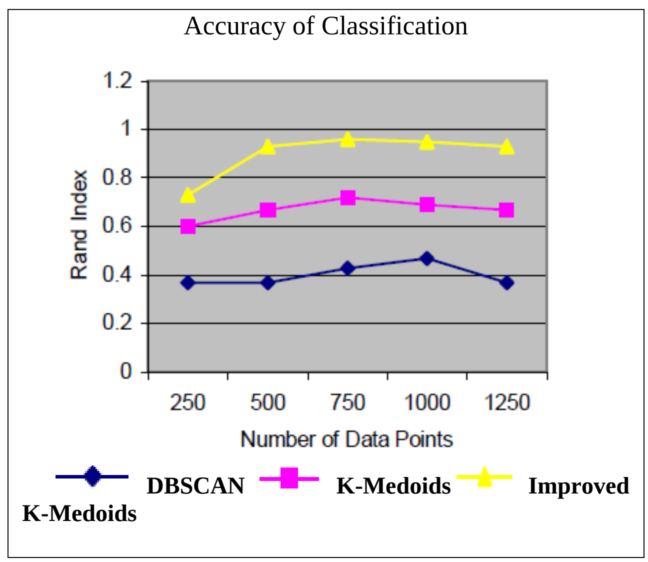
The Number Of Classes: 5

The Number Of Dimensions: 2

The Number Of Points Per Class: 50 , 100, 150, 200,250

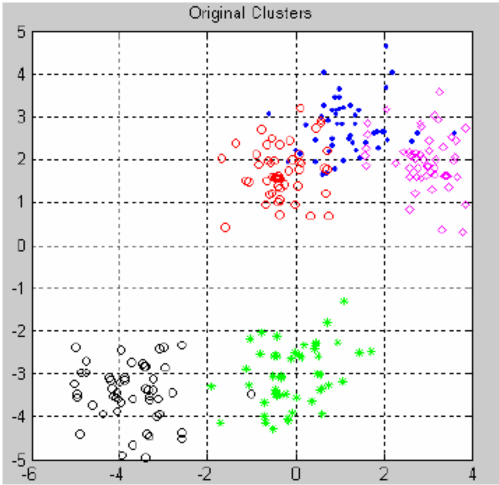
The standard Deviation: 7.000000e-001

S.No.	Total No. of Records in Proteins Data Base	Classification Accuracy Measured by Rand Index (Found using Original Class Labels and Calculated Class Labels)		
		DBSCAN	K-Medoids	Density based Improved K-Medoids
1	250	0.03	0.02	0.03
2	500	0.09	0.02	0.09
3	750	0.2	0.02	0.2
4	1000	0.31	0.03	0.3
5	1250	0.42	0.01	0.44

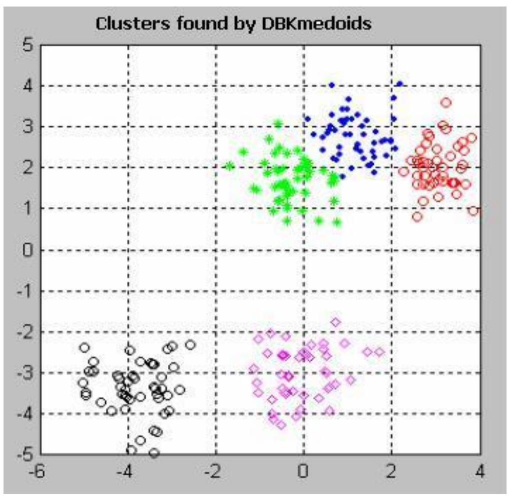


The Clustering and Outlier Detection Results

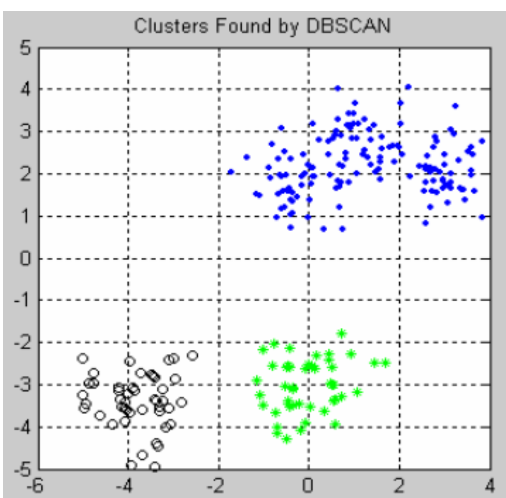
The following show the clusters and outliers marked with DBSCAN and DBkmedoids



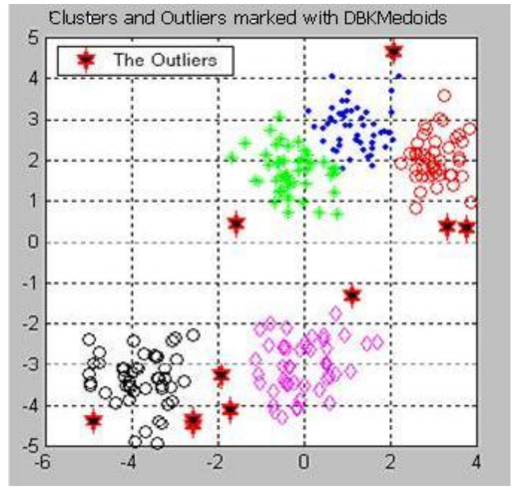
Original Custers



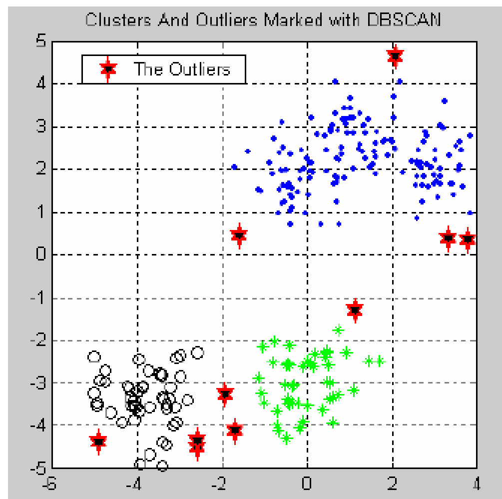
Clusters found by DBKmedoids



Clusters found by DBSCAN



Clusters and outliers marked with DBKMedoids



Clusters and Outliers marked with DBSCAN

From the plotted results, it is noted that DBkmedoids perform better than DBSCAN.

V. CONCLUSION

This Clustering is an efficient way of reaching information from raw data and Kmeans, Kmedoids are basic methods for it. Although it is easy to implement and understand, Kmeans and Kmedoids have serious drawbacks. The proposed clustering and outlier detection system has been implemented using Weka and tested with the proteins data base created by Gaussian distribution function. The data will form circular or spherical clusters in space. As shown in the tables and graphs, the proposed Density based Kmedoids algorithm performed very well than DBSCAN and k-medoids clustering in term of quality of classification measured by Rand index. One of the major challenges in medical domain is the extraction of comprehensible knowledge from medical diagnosis data.

There is lot of scope for the proposed Density based KMedoidss clustering algorithm in different application areas such as medical image segmentation and medical data mining. Future works may address the issues involved in applying the algorithm in a particular application area.

References

[1] Shu-Chuan Chu;Roddick, J.F.;Tsong-Yi Chen;Jeng-Shyang Pan "Efficient search approaches for k-medoids-based algorithms", TENCON '02. Proceedings. 2002 IEEE Region 10 Conference on Computers, Communications, Control and Power Engineering,Feb 2003.

[2] Shu-Chuan Chu,John F.Roddick,J.S. Pan "An Efficient K -Medoids-Based Algorithm Using Previous Medoid Index, Triangular Inequality Elimination Criteria, and Partial Distance Search" 4th International Conference on Data Warehousing and Knowledge Discovery,2002.

[3] Raymond T. Ng and Jiawei Han, CLARANS: A Method for Clustering Objects for Spatial Data Mining, IEEE TRANSACTIONS ON KNOWLEDGE and DATA ENGINEERING, Vol. 14, No. 5, SEPTEMBER 2002.

[4] K. Mumtaz1 and Dr. K. Duraiswamy, A Novel Density based improved k-means Clustering Algorithm – Dbkmeans International Journal on Computer Science and Engineering, Vol. 02, No. 02, 2010, 213-218.

[5] Zhang T, Ramakrishnan R., Livny M., and BIRCH: An efficient data clustering method for very large databases, In: SIGMOD Conference, pp.103~114, 1996.

[6] K. Alsabti, S. Ranka, and V. Singh, ªAn Efficient k-means Clustering Algorithm,º Proc. First Workshop High Performance Data Mining, Mar. 1998.

[7] M. Ester, H. P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases", Proc. 1996 Int. Conf. Knowledge Discovery and Data Mining (KDD'96), Portland, OR, August 1996.

[8] Matheus C.J.; Chan P.K.; and Piatetsky-Shapiro G. 1993. Systems for Knowledge Discovery in Databases, IEEE Transactions on Knowledge and Data Engineering 5(6): 903-913.

[9] Ng R.T. and Han J. 1994. Efficient and Effective Clustering Methods for Spatial Data Mining, Proc. 20th Int. Conf. on Very Large Data Bases, 144-155. Santiago, Chile.

[10] Indukuri, R. K. R. (2011). Dominating Sets and Spanning Tree based Clustering Algorithms for Mobile Ad hoc Networks. IJACSA - International Journal of Advanced Computer Science and Applications, 2(2). Retrieved from http://ijacsa.thesai.org.

[11] Kumar, V. (2011). Knowledge discovery from database Using an integration of clustering and classification. IJACSA - International Journal of Advanced Computer Science and Applications, 2(3), 29-33. Retrieved from http://ijacsa.thesai.org.

[12] Jayech, K. (2011). Clustering and Bayesian network for image of faces classification. IJACSA - International Journal of Advanced Computer Science and Applications, (Special Issue). Retrieved from www.ijacsa.thesai.org.

Context Switching Semaphore with Data Security Issues using Self-healing Approach

M. Anand
Research Scholar, Dept of ECE
St. Peters University
Chennai - 54 India

Dr. S. Ravi
Professor and Head, Dept of ECE
Dr M. G. R. University
Chennai – 95 India

Kuldeep Chouhan
Research Scholar, Dept of ECE
Dr M. G. R. University
Chennai – 95 India

Syed Musthak Ahmed
Head, Dept of ECE
S. R. Engineering College
Warangal India

***Abstract*— The main objective of a self healing scheme is to share and secure the information of any system at the same time. "Self-healing" techniques ultimately are dependable computing techniques. Specifically self-healing systems have to think for itself without human input, able to boot up backup systems. However, sharing and protection are two contradictory goals. Protection programs may be completely isolated from each other by executing them on separate non-networked computer, however, this precludes sharing.**

Keywords- Self-healing; Semaphore; Data Security.

I. INTRODUCTION

Self-healing mechanisms complement approaches that stop attacks from succeeding by preventing the injection of code, transfer of control to injected code, or misuse of existing code. Approaches to automatically defending software systems have typically focused on ways to proactively or at runtime protect an application from attack. These proactive approaches include writing the system in a "safe" language, linking the system with "safe" libraries, transforming the program with artificial diversity, or compiling the program with stack integrity checking. The technique of program shepherding is validates branch instructions to prevent transfer of control to injected code and to make sure that calls into native libraries originate from valid sources. Control Flow Integrity (CFI), observing that high-level programming often assumes properties of control flow that is not enforced at the machine level [3,4]. The use of CFI enables the efficient implementation of a software shadow call stack with strong protection guarantees. However, such techniques generally focus on integrity protection at the expense of availability. Control flow is often corrupted because input is eventually incorporated into part of an instruction's opcode, set as a jump target, or forms part of an argument to a sensitive system call.

II. SELF HEALING SYSTEMS

A. Self Healing Approach

Self-healing is an approach to detect improper operations of software applications, transactions and business processes, and then to initiate corrective action without disrupting users [8]. Healing systems that require human intervention or intervention of an agent external to the system can be categorized as assisted-healing systems. The key focus or contrasting idea as compared to dependable systems is that a self-healing system should recover from the abnormal (or unhealthy) state and return to the normative (healthy) state and function as it was prior to disruption. Some scholars treat self-healing systems as an independent one while others view as a subclass of traditional fault tolerant computing systems.

The system monitors itself for indications of anomalous behavior. When such behavior is detected, the system enters a self-diagnosis mode that aims to identify the fault and extract as much information as possible with respect to its cause, symptoms, and impact on the system [9]. The system tries to adapt itself by generating candidate fixes, which are tested to find the best target state. Self-healing systems can support decision making in a large way for managerial and organizational situations [11]. Many of the decision support systems (DSS) offer passive forms of decision support, where the decision-making process depends upon the user's initiative. Such active involvement is especially needed in complex decision-making environments.

B. Architecture of Self-Healing (S-H)

The term 'self' in self-healing architecture is referred to the action or response initiated automatically within the system. A general architecture of a self-healing system is shown in Fig. 1.

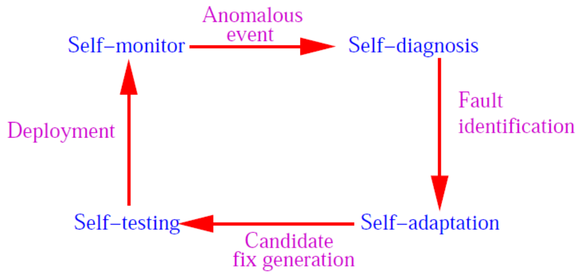


Fig. 1 General architecture of a self-healing system

C. *Self Healing technique*

The effective remediation strategies include failure-oblivious computing, error virtualization, rollback of memory updates, and data-structure repair [5]. These approaches may cause a semantically incorrect continuation of execution attempts to address this difficulty by exploring semantically safe alterations of the program's environment. The technique is subsequently introduced in a modified form as failure-oblivious computing, because the program code is extensively rewritten to include the necessary checks for every memory access, the system incurs overheads for a variety of different applications. Data-structure Repair is the most critical concerns with recovering from software faults and vulnerability exploits is ensuring the consistency and correctness of program data and state.

D. *Why Self Healing Systems*

The software notoriously buggy and crash-prone is despite considerable work in the fault tolerance and reliability [1]. The current approach to ensuring the security and availability of software consists of a mix of different techniques:

***a) Proactive techniques*:** seek to make the code as dependable as possible, through a combination of safe languages, libraries and compilers, code analysis tools, formal methods and development methodologies.

b) *Debugging techniques*: aim to make post-fault analysis and recovery as easy as possible for the programmer that is responsible for producing a fix.

c) *Runtime protection techniques*: try to detect the fault using some type of fault isolation, which address specific types of faults or security vulnerabilities.

d) *Containment techniques*: seek to minimize the scope of a successful exploit by isolating the process from the rest of the system, *e.g.*, through use of virtual machine.

e) *Byzantine fault-tolerance and quorum techniques*: rely on redundancy and diversity to create reliable systems out of unreliable components.

E. *Elements of Self-Healing model*

In the Self-healing process model, there are different categories of aspects to the self-healing system,

a) ***Fault model*:** Self-healing systems have the tenets of dependable computing is that called a fault model must be specified for any fault tolerant system. The fault model answers the question of what faults the system is to tolerate. Self-healing systems have a fault model in terms of what injuries (faults), which are expected to be able to self-heal.

b) ***Fault duration*:** Faults can be permanent, intermittent or transient due to an environmental condition. It is important to state the fault duration assumption of a self-healing approach to understand what situations it addresses.

c) ***Fault manifestation*:** The severity of the fault manifestation, it affects the system in the absence of a self-healing response. The faults cause immediate system crashes, but, many faults cause less catastrophic consequences, such as system slow-down due to excessive CPU loads, thrashing due to memory hierarchy overloads, resource leakage, file system overflow, and so on.

d) ***Fault source*:** Thee source of faults can affect self-healing strategies due to implementation defects, requirements defects, operational mistakes, and etc. Self-healing software is designed only to withstand hardware failures such as loss of memory [6] or CPU capacity and not software failures.

e) ***Granularity*:** The granularity of a failure is the size of the component that is compromised by that fault. Different self-healing mechanisms are probably appropriate depending on the granularity of the failures [7] and hence, the granularity of recovery actions.

f) ***Fault profile expectations*:** The source of the fault is the profile of fault occurrences that is expected. It considered for self-healing might be only expected faults that is based on design analysis or faults that are unexpected [12]. Additionally, faults might be random and independent, might be correlated in space or time, or might even be intentional due to malicious intent.

F. *Conventional Methods of Security*

The conventional methods can overcome only the effects of passive threats and not the active threats for the authenticate users. They reduce user-friendliness and also, the amount of OS resources required to provide security is high. Different protocol architectures are used for providing security for each layer of the OSI model and it may not be generic. Alternately, in this work, the information is allowed to flow freely through the fetch and decode cycles while an access or authentication is made only between the decode and execute cycle before the data is permanently written into the memory by the user (if authenticated). This is shown in Fig. 2.

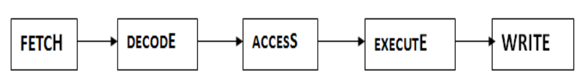


Fig. 2 Process of Execution

G. *Proposed features of Security issues*

The proposed hardware is shown in Figure 3. The features of the proposed hardware are that it is (i) PCI compliant and (ii) Mounted in a single chip

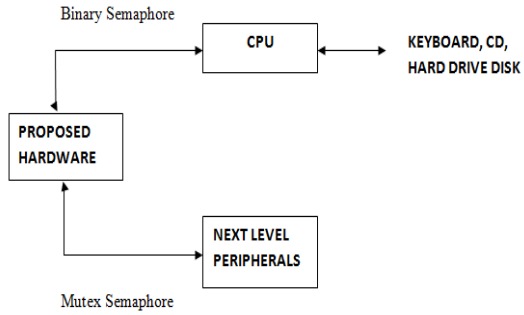


Fig. 3 Hardware Process System

a) ***Robustness***: It provides defence against vulnerabilities with few false positives or false negatives.

b) ***Flexible***: It adapts easily to cover the continuously evolving threats.

c) ***End-to-End***: The security policy flows throughout all the seven layers of the OSI model.

d) ***Scalable***: It can co-exist with the existing circuitry without any modification [2].

III. METHODOLOGY

A. Semaphores and Resource Sharing

A semaphore is a protected variable or an abstract data type which restricts the access to shared resources such as shared memory in a multiprogramming environment. It is a primitive synchronization mechanism for sharing CPU time and resources. It is a classic solution to prevent race conditions.

B. Operation of Semaphore

The 'value' of a semaphore is the number of units of resources which are free. To avoid busy-waiting, a semaphore has an associated queue of processes (usually First in First Out). If a process performs a 'P' operation on a semaphore which has the value zero, the process is added to the semaphore's queue. When another process increments the semaphore by performing a 'V' operation, and there are processes on the queue, one of them is removed from the queue and resumes operation.

C. Binary Semaphore

In binary semaphore, if there is only one resource, the semaphore takes value '0' or '1'.This is explained in Fig. 4.

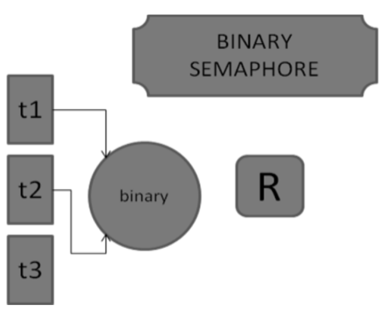


Fig. 4 Binary Semaphore

Suppose a 'P' operation busy-waits (uses its turn to do nothing) or maybe sleeps (tells the system not to give it a turn) until a resource is available, where upon it immediately claims one. Now, let 'V' be the operation that simply makes a resource available again after the process has finished using it. The 'P' and 'V' operations must be atomic, i.e., no process may be preempted in the middle of one of those operations to run another operation on the same semaphore. When a semaphore is being used, it takes value '0' and when it takes the value '1', the process directly starts execution without waiting.

D. Counting Semaphore

The counting semaphore concept can be extended with the ability of claiming or returning more than one unit from the semaphore. When multiple resources are to be shared by many operations, such a semaphore is used. All the resources must be of the same type. This is shown in Fig. 5.

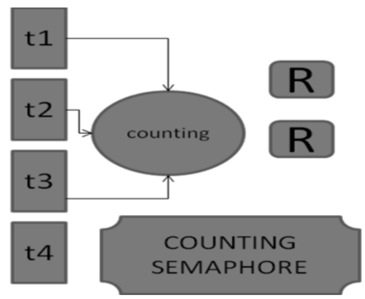


Fig. 5 Counting Semaphore

In this, the initial value of semaphore is set equal to number of resources available. As the semaphores are being used, the value keeps decrementing. As the semaphores are being released, after use, its value keeps incrementing. 'Zero' value refers to empty semaphore.

E. Mutex Semaphore

A mutex is a binary semaphore with extra features like ownership or priority inversion protection. Mutexes are meant to be used for mutual exclusion only. This is shown in Figure 6.

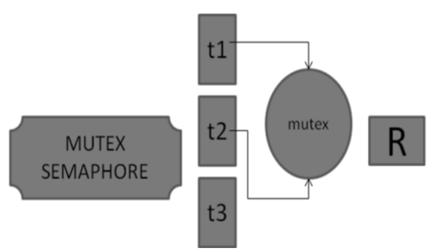


Fig. 6 Mutex Semaphore

Initially the semaphore value is set to zero. Once a task attains ownership, it can access the resources as many times as it wants and each time, it accesses, the semaphore value increases.

F. Characteristics of Semaphore

The characteristics of the three semaphores are shown in TABLE I.

TABLE I. CHARACTERISTICS OF THREE SEMAPHORES

Binary Semaphore	Counting Semaphore	Mutex Semaphore
Anyone can release the semaphore. Used for mutual exclusion and event notification.	Only after a task has attained access, it can release the semaphore.	Only the owner can release the semaphore. Used only for mutual exclusion.

IV. RESULTS AND DISCUSSION

A. Implementation

In the existing architecture of computers, the bottleneck of connecting a high speed low memory device to a low speed high memory device is solved by using an intermediate memory module called the cache. The cache exists between the high speed CPU and the lower speed memories. However, there is no security between the data link layer of the CPU and the cache. The proposed card is placed on the PCI bus at the maximum possible speed and a direct connection is established to the CPU via snooping. To understand how the security layer is to be implemented, the knowledge of the three basic terms is required: subject, object and capability. Subject refers to the user or entity which acts on behalf of the user on the system. Objects may be defined as resources within the system. The main term however is capability which is basically a 'token'. The possession of a capability by a subject confers access right for an object. They cannot be easily modified, but they can be reproduced.

The capabilities of the objects are to be stored in the non-readable section of the HDD. For a subject to access a particular object, it must possess the capability for doing so. Hence, before a subject accesses a resource via the CPU, it will first go through a screening check from the hardware on whether or not its capabilities allow it to access such resources. Hence, a security layer is now added to the data link layer between the CPU and the cache. Additionally, user also must be prevented from creating arbitrary capabilities. This can be accomplished by placing the capabilities in special 'Capability Segments' which users cannot access. Another approach is to add a tag bit to each primary storage location. This bit, inaccessible to the user is 'ON' if the location contains a capability. It should be noted that the hardware restricts the manipulation of the location contents to appropriate system routines. If the last remaining capability is destroyed, then that object cannot be used in any manner. In this work, special provisions are made for controlling the copying and movement of capabilities (as well as interpretation) depending on the hardware involved.

B. Evolved Function is a Semaphore Selector

The schematic of the control block performing the evolved function is shown in Fig. 7.

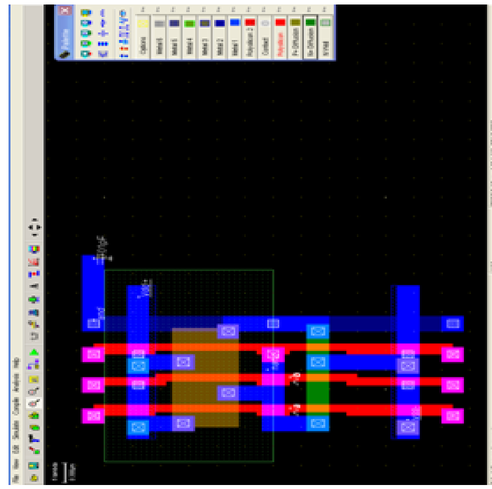


Fig. 7 Schematic representation of Semaphore Selector

The power consumed by the block under read and write mode is shown in Fig. 8 and Fig. 9 respectively.

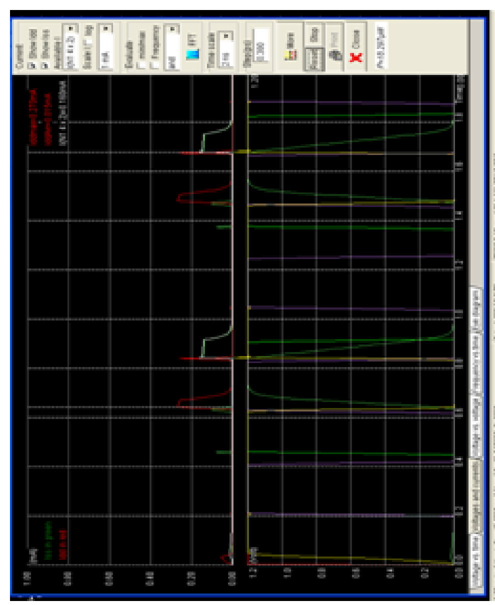


Fig. 8 Current and Voltage Variations of semaphore selector during read cycle

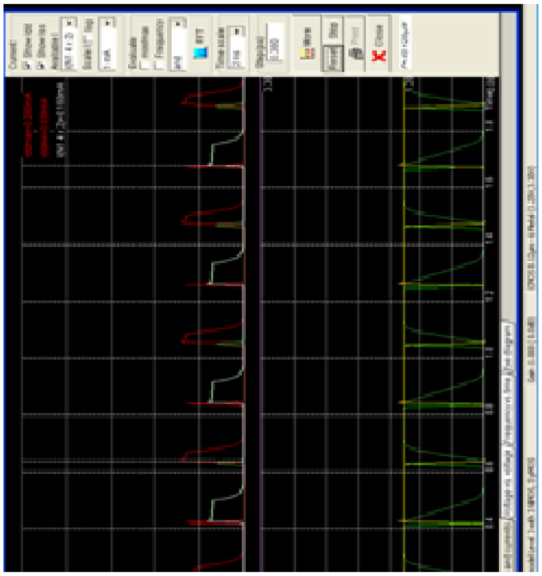


Fig. 9 Current and Voltage Variations of Semaphore selector during write cycle

C. Evolved Function is a Resource Sharing Selector

The control block performing the evolved function of “resource sharing selector” along with its power consumtion is shown in Fig. 10. Similar graph corresponding to read and write cycle is shown in Fig. 11 and Fig. 12 respectively.

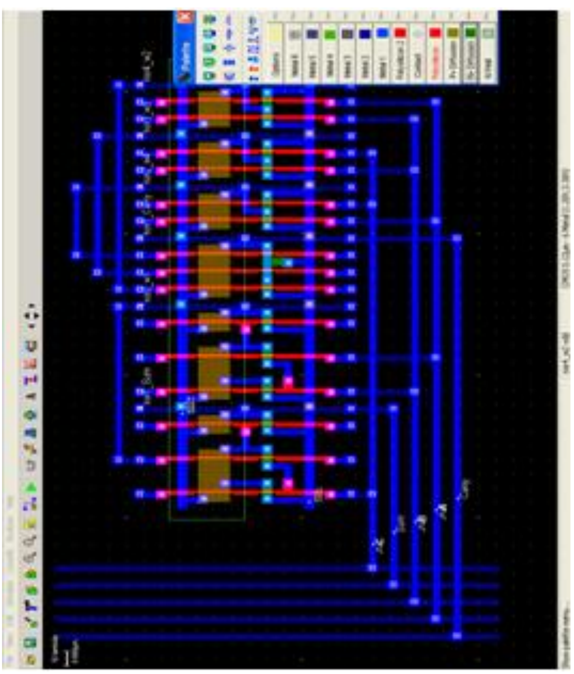


Fig. 10 Schematic of Evolved resource sharing selector

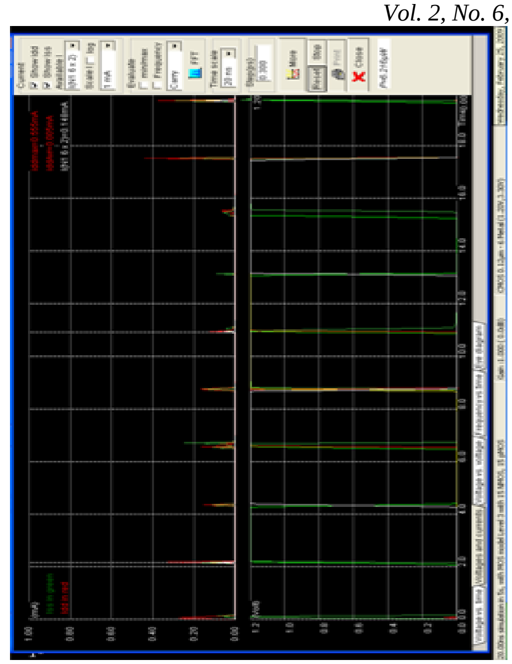


Fig. 11 Current and Voltage Variations of evolved resource sharing selector during Read cycle

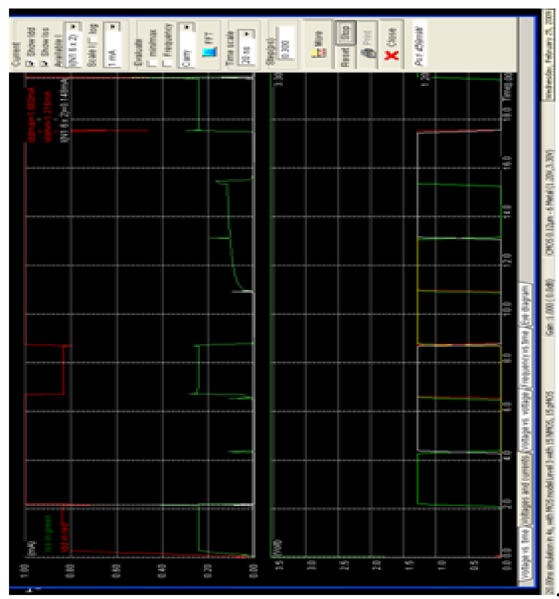


Fig. 12 Current and Voltage Variations of evolved resource sharing selector during write cycle

D. Evolved Function is a Snoop Selector Function

The schematic of the control block performing the evolved function of a snoop selector circuit is shown in Fig. 13. The power consumed by the PE under read and write cycles is shown in Fig. 14 and Fig. 15 respectively.

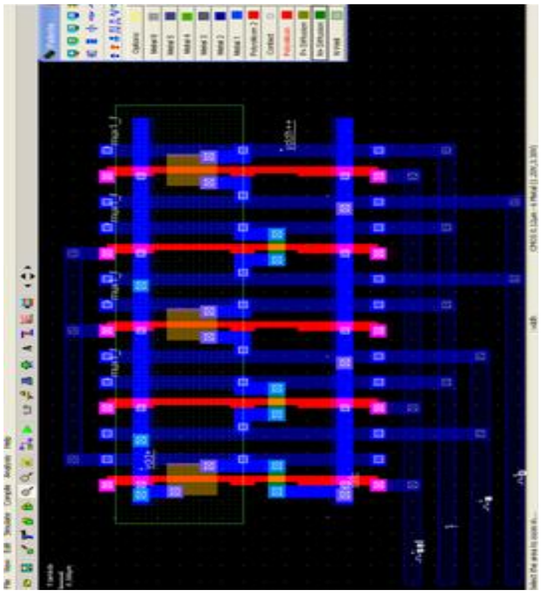


Fig. 13 Schematic of Evolved Snoop selector circuit

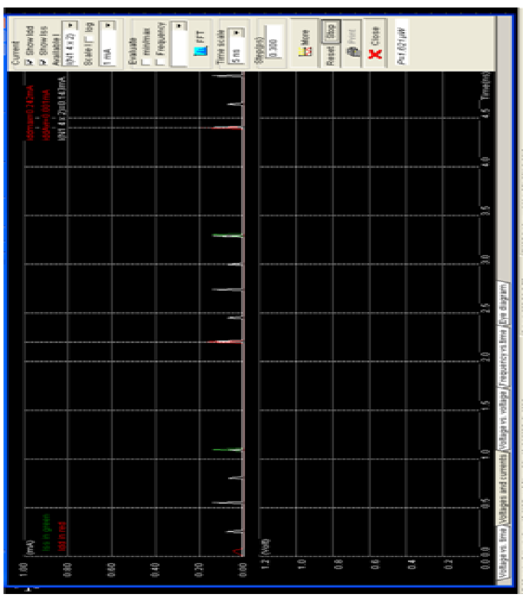


Fig. 14 Current and Voltage Variations of evolved Snoop selector during read
Cycle

E. Evolved Function is a Context Switching Semaphore

The schematic of Context switching semaphore block of the evolved function is shown in Fig. 16. The power consumed by the PE during read and write cycle is shown in Fig. 17 and Fig. 18 respectively

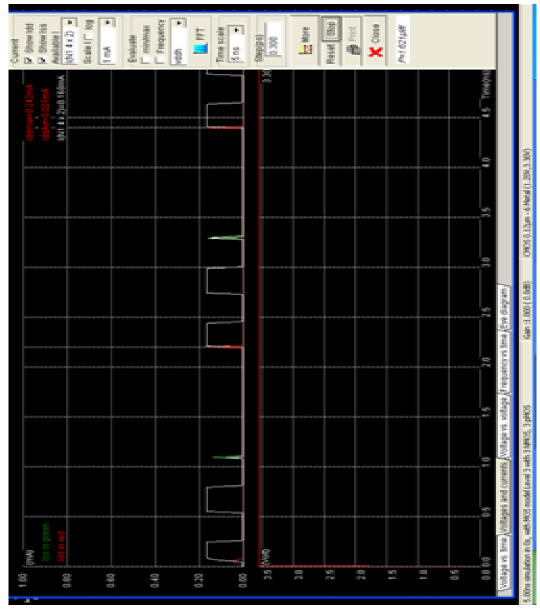


Fig. 15 Current and Voltage Variations of evolved snoop selector during write
cycle

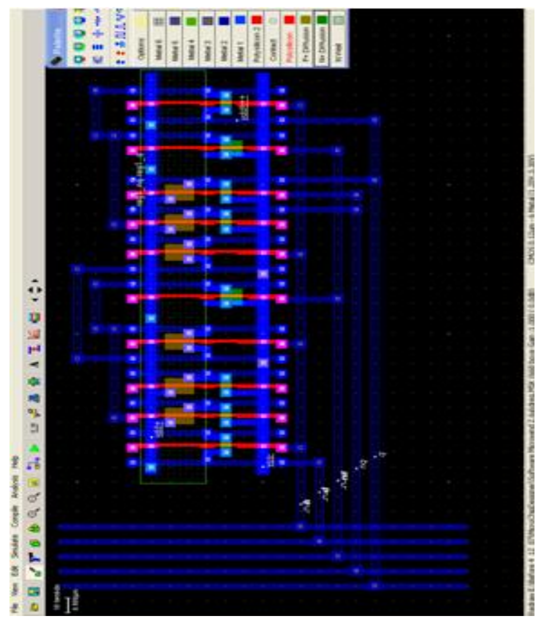


Fig. 16 Schematic of Evolved Context switching semaphore

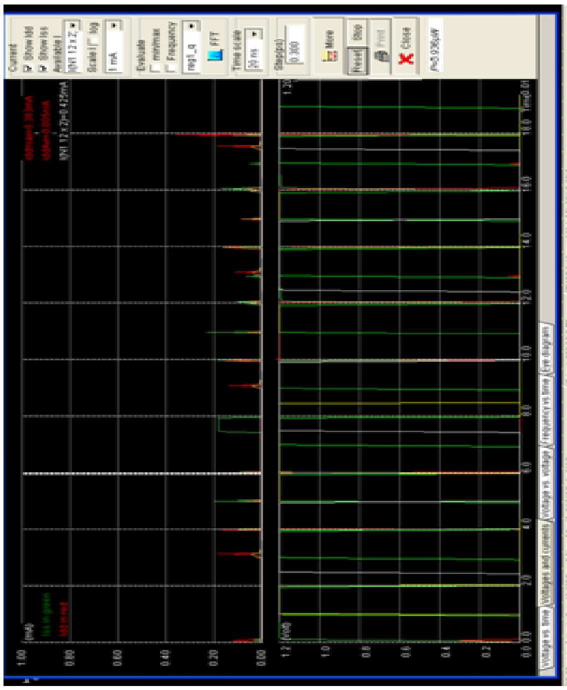


Fig. 17 Current and Voltage Variations of evolved Context switching semaphore during read cycle

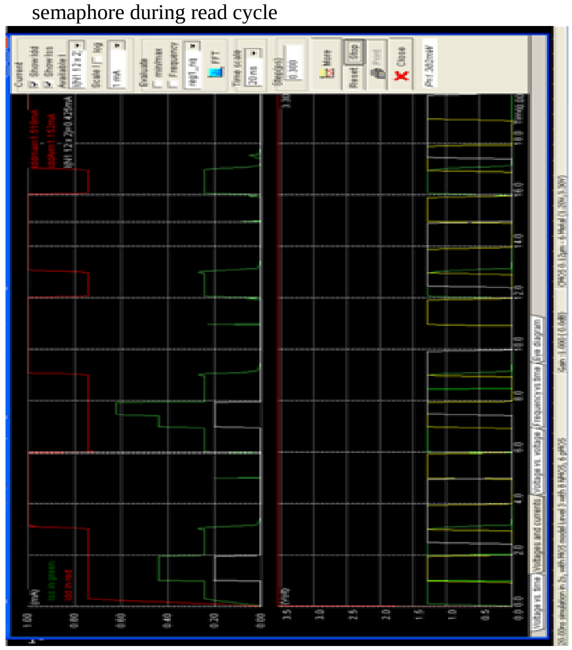


Fig. 18 Current and Voltage Variations of evolved Context switching semaphore during write cycle

The results of the discussion are tabulated in TABLE II. It can be inferred from the table that a power level variation exists between the read and write cycles.

TABLE II. POWER CONSUMED BY EVOLVED PE

Evolved Function of PE	**Average Power Consumed by Evolved blocks**	
	read	**write**
Semaphore selector	.3mW	2.325mW
Resource sharing selector	.54mW	3.5mW
Snoop selector	.12mW	3.665mW
Context switching semaphore	.456mW	3.5mW

V. CONCLUSION

In this work, self-healing systems prove increasingly important in countering system software based attacks, which recover and secure to the data from interrupted services. Self-healing systems offer an active form of decision support, without human intervention that can detect the fault and recover from the fault. Also, with intelligent architectural models, a self-healing system can select the proper repair plan to deploy the broken component, if there is more than one component that needs to be healed, can prioritize a fault component over the others, etc.

ACKNOWLEDGMENT

This work was supported in part by Micro Logic System, Chennai-600017.

REFERENCES

[1] Bouricius, W.G., Carter, W.C. & Schneider, P.R, "Reliability modeling techniques for self-repairing computer systems", in proceedings of 24th National Conference, ACM, 1969, pp. 395-309.

[2] Shelton, C., Koopman, P.&Nace, W., "A framework for scalable analysis and design of system-wide graceful degradation in distributed embedded systems", WORDS03, January 2003.

[3] G. Edward Suh, J. W. Lee, D. Zhang, and S. Devadas, "Secure Program Execution via Dynamic Information Flow Tracking", in proceedings of the 11th International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS-XI), October 2004.

[4] Martin Abadi, Mihai Budiu, Ulfar Erlingsson, and Jay Ligatti, "Control-Flow Integrity: Principles, Implementations, and Applications", in proceedings of the ACM Conference on Computer and Communications Security (CCS), 2005.

[5] B. Demsky and M. C. Rinard, "Automatic Detection and Repair of Errors in Data Structures", in proceedings of ACM OOPSLA, October 2003.

[6] M. Locasto, S. Sidiroglou, and A.D Keromytis, "Software Self-Healing Using Collaborative Application Communities", in proceedings of the Internet Society (ISOC) Symposium on Network and Distributed Systems Security (SNDSS), February 2006.

[7] J. Kong, X. Hong, J.-S. Park, Y. Yi, and M. Gerla, "L'Hospital: Self-healing Secure Routing for Mobile Ad-hoc Networks", in Technical Report CSD-TR040055, Dept. of Computer Science, UCLA, January 2005.

[8] Michael E. Shin and Jung Hoon An, "Self-Reconfiguration in Self-Healing Systems", in proceedings of the Third IEEE International Workshop on EASE'06, pp 89-98 (2006).

[9] E. M. Dashofy, A. V. D. Hoek, and R. N. Taylor, "Towards architecture-based self-healing systems", in p*roceedings of the first workshop on Self-healing systems*, Charleston, South Carolina, pp. 21-26, 2002.

[10] H.You, V.Vittal, Z.Yang, "Self-healing in power systems: an approach using islanding and rate of frequency decline based load shedding", IEEE Transaction on Power System, Vol.18, No.1, 2003, pp.174 -181.

[11] T.A. Ramesh Kumaar and Dr.I.A.Chidambaram, "Self-Healing Strategy for Dynamic Security Assessment and Power System

Restoration", in International Journal of Computer Science & Emerging Technologies, (E-ISSN: 2044-6004) Vol. 2, Issue 2, April 2011.

[12] J. Newsome and D. X. Song, "Dynamic Taint Analysis, and Signature Generation of Exploits on Commodity Software", in the Proceedings of the Network and Distributed System Security Symposium, San Diego, CA, Feb.2005

AUTHORS PROFILE

1. Mr. M. Anand has completed his B.Tech. in the faculty of ECE and M.Tech. in the area of Applied Electronics. He is presently a research scholar in the faculty of ECE at St. Peters University, Avadi, Chennai. Mr. M.Anand shows interest in the areas of Embedded system design, Co-design of embedded systems, Scheduling etc.

2. Dr. S. Ravi was born on July 24th, 1971 at Chennai, India and got A.M.I.E. (Electronics Engineering) from Institution of Engineers, Calcutta, M.Tech. (Communication systems) from Anna University, Chennai in the Year 1994 and Ph.D. from Anna University in the year 2003. He has got 20 Years of Teaching Experience and 3 Years of Industrial Experience. Presently, he is working as Professor and Head of the Department of Electronics Engineering in Dr. M.G.R. University, Chennai. He has so far published nearly thirty four papers in referred International Journals and successfully guided four Ph.D. scholars. Dr. S. Ravi is a Life Member IETE and Life Member of Indian Society for Technical Education

3. Mr. Kuldeep Chouhan has completed his M.Tech. in the department of Computer Science and Engineering (CSE). He is presently a research scholar in the faculty of ECE at Dr. MGR University, Chennai. His area of interest is Wireless Sensor Network, Embedded system, MultiAgent System in Sensor Security

4. Dr. Syed Musthak Ahmed, completed his BE(Electronics) from Bangalore University, ME(Electronics), UVCE Bangalore, Bangalore University, Ph.D.(ECE), Vinayaka Missions University, Salem. He has 25 years of teaching experience working in reputed engineering colleges and is presently working as Professor and HOD (ECE), SR Engineering College, Warangal. He is a member of Various Professional Societies like SMIEEE, FIETE, MISSS, MISTE, MIAENG, MIAMT. He has various publications in National and International Journals / Conferences

Mining Educational Data to Analyze Students' Performance

Brijesh Kumar Baradwaj
Research Scholor,
Singhaniya University,
Rajasthan, India

Saurabh Pal
Sr. Lecturer, Dept. of MCA,
VBS Purvanchal University,
Jaunpur-222001, India

***Abstract*— The main objective of higher education institutions is to provide quality education to its students. One way to achieve highest level of quality in higher education system is by discovering knowledge for prediction regarding enrolment of students in a particular course, alienation of traditional classroom teaching model, detection of unfair means used in online examination, detection of abnormal values in the result sheets of the students, prediction about students' performance and so on. The knowledge is hidden among the educational data set and it is extractable through data mining techniques. Present paper is designed to justify the capabilities of data mining techniques in context of higher education by offering a data mining model for higher education system in the university. In this research, the classification task is used to evaluate student's performance and as there are many approaches that are used for data classification, the decision tree method is used here.**

By this task we extract knowledge that describes students' performance in end semester examination. It helps earlier in identifying the dropouts and students who need special attention and allow the teacher to provide appropriate advising/counseling.

Keywords-Educational Data Mining (EDM); Classification; Knowledge Discovery in Database (KDD); ID3 Algorithm.

I. INTRODUCTION

The advent of information technology in various fields has lead the large volumes of data storage in various formats like records, files, documents, images, sound, videos, scientific data and many new data formats. The data collected from different applications require proper method of extracting knowledge from large repositories for better decision making. Knowledge discovery in databases (KDD), often called data mining, aims at the discovery of useful information from large collections of data [1]. The main functions of data mining are applying various methods and algorithms in order to discover and extract patterns of stored data [2]. Data mining and knowledge discovery applications have got a rich focus due to its significance in decision making and it has become an essential component in various organizations. Data mining techniques have been introduced into new fields of Statistics, Databases, Machine Learning, Pattern Reorganization, Artificial Intelligence and Computation capabilities etc.

There are increasing research interests in using data mining in education. This new emerging field, called Educational Data Mining, concerns with developing methods that discover knowledge from data originating from educational environments [3]. Educational Data Mining uses many techniques such as Decision Trees, Neural Networks, Naïve Bayes, K- Nearest neighbor, and many others.

Using these techniques many kinds of knowledge can be discovered such as association rules, classifications and clustering. The discovered knowledge can be used for prediction regarding enrolment of students in a particular course, alienation of traditional classroom teaching model, detection of unfair means used in online examination, detection of abnormal values in the result sheets of the students, prediction about students' performance and so on.

The main objective of this paper is to use data mining methodologies to study students' performance in the courses. Data mining provides many tasks that could be used to study the student performance. In this research, the classification task is used to evaluate student's performance and as there are many approaches that are used for data classification, the decision tree method is used here. Information's like Attendance, Class test, Seminar and Assignment marks were collected from the student's management system, to predict the performance at the end of the semester. This paper investigates the accuracy of Decision tree techniques for predicting student performance.

II. DATA MINING DEFINITION AND TECHNIQUES

Data mining, also popularly known as Knowledge Discovery in Database, refers to extracting or "mining" knowledge from large amounts of data. Data mining techniques are used to operate on large volumes of data to discover hidden patterns and relationships helpful in decision making. While data mining and knowledge discovery in database are frequently treated as synonyms, data mining is actually part of the knowledge discovery process. The sequences of steps identified in extracting knowledge from data are shown in Figure 1.

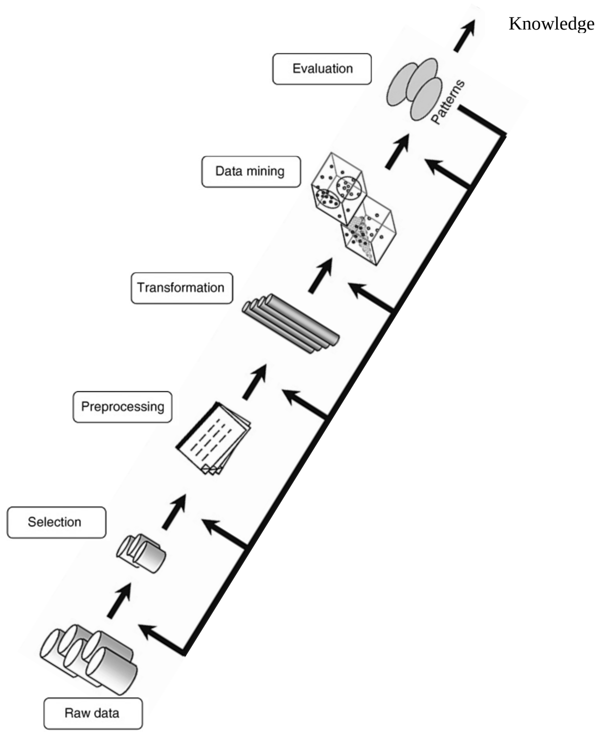


Figure 1: The steps of extracting knowledge from data

Various algorithms and techniques like Classification, Clustering, Regression, Artificial Intelligence, Neural Networks, Association Rules, Decision Trees, Genetic Algorithm, Nearest Neighbor method etc., are used for knowledge discovery from databases. These techniques and methods in data mining need brief mention to have better understanding.

A. Classification

Classification is the most commonly applied data mining technique, which employs a set of pre-classified examples to develop a model that can classify the population of records at large. This approach frequently employs decision tree or neural network-based classification algorithms. The data classification process involves learning and classification. In Learning the training data are analyzed by classification algorithm. In classification test data are used to estimate the accuracy of the classification rules. If the accuracy is acceptable the rules can be applied to the new data tuples. The classifier-training algorithm uses these pre-classified examples to determine the set of parameters required for proper discrimination. The algorithm then encodes these parameters into a model called a classifier.

B. Clustering

Clustering can be said as identification of similar classes of objects. By using clustering techniques we can further identify dense and sparse regions in object space and can discover overall distribution pattern and correlations among data attributes. Classification approach can also be used for effective means of distinguishing groups or classes of object but it becomes costly so clustering can be used as preprocessing approach for attribute subset selection and classification.

C. Predication

Regression technique can be adapted for predication. Regression analysis can be used to model the relationship between one or more independent variables and dependent variables. In data mining independent variables are attributes already known and response variables are what we want to predict. Unfortunately, many real-world problems are not simply prediction. Therefore, more complex techniques (e.g., logistic regression, decision trees, or neural nets) may be necessary to forecast future values. The same model types can often be used for both regression and classification. For example, the CART (Classification and Regression Trees) decision tree algorithm can be used to build both classification trees (to classify categorical response variables) and regression trees (to forecast continuous response variables). Neural networks too can create both classification and regression models.

D. Association rule

Association and correlation is usually to find frequent item set findings among large data sets. This type of finding helps businesses to make certain decisions, such as catalogue design, cross marketing and customer shopping behavior analysis. Association Rule algorithms need to be able to generate rules with confidence values less than one. However the number of possible Association Rules for a given dataset is generally very large and a high proportion of the rules are usually of little (if any) value.

E. Neural networks

Neural network is a set of connected input/output units and each connection has a weight present with it. During the learning phase, network learns by adjusting weights so as to be able to predict the correct class labels of the input tuples. Neural networks have the remarkable ability to derive meaning from complicated or imprecise data and can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. These are well suited for continuous valued inputs and outputs. Neural networks are best at identifying patterns or trends in data and well suited for prediction or forecasting needs.

F. Decision Trees

Decision tree is tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. Specific decision tree methods include Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID).

G. Nearest Neighbor Method

A technique that classifies each record in a dataset based on a combination of the classes of the k record(s) most similar to it in a historical dataset (where k is greater than or equal to 1). Sometimes called the k-nearest neighbor technique.

III. Related Work

Data mining in higher education is a recent research field and this area of research is gaining popularity because of its potentials to educational institutes.

Data Mining can be used in educational field to enhance our understanding of learning process to focus on identifying, extracting and evaluating variables related to the learning process of students as described by Alaa el-Halees [4]. Mining in educational environment is called Educational Data Mining.

Han and Kamber [3] describes data mining software that allow the users to analyze data from different dimensions, categorize it and summarize the relationships which are identified during the mining process.

Pandey and Pal [5] conducted study on the student performance based by selecting 600 students from different colleges of Dr. R. M. L. Awadh University, Faizabad, India. By means of Bayes Classification on category, language and background qualification, it was found that whether new comer students will performer or not.

Hijazi and Naqvi [6] conducted as study on the student performance by selecting a sample of 300 students (225 males, 75 females) from a group of colleges affiliated to Punjab university of Pakistan. The hypothesis that was stated as "Student's attitude towards attendance in class, hours spent in study on daily basis after college, students' family income, students' mother's age and mother's education are significantly related with student performance" was framed. By means of simple linear regression analysis, it was found that the factors like mother's education and student's family income were highly correlated with the student academic performance.

Khan [7] conducted a performance study on 400 students comprising 200 boys and 200 girls selected from the senior secondary school of Aligarh Muslim University, Aligarh, India with a main objective to establish the prognostic value of different measures of cognition, personality and demographic variables for success at higher secondary level in science stream. The selection was based on cluster sampling technique in which the entire population of interest was divided into groups, or clusters, and a random sample of these clusters was selected for further analyses. It was found that girls with high socio-economic status had relatively higher academic achievement in science stream and boys with low socio-economic status had relatively higher academic achievement in general.

Galit [8] gave a case study that use students data to analyze their learning behavior to predict the results and to warn students at risk before their final exams.

Al-Radaideh, et al [9] applied a decision tree model to predict the final grade of students who studied the C++ course in Yarmouk University, Jordan in the year 2005. Three different classification methods namely ID3, C4.5, and the NaïveBayes were used. The outcome of their results indicated that Decision Tree model had better prediction than other models.

Pandey and Pal [10] conducted study on the student performance based by selecting 60 students from a degree college of Dr. R. M. L. Awadh University, Faizabad, India. By means of association rule they find the interestingness of student in opting class teaching language.

Ayesha, Mustafa, Sattar and Khan [11] describes the use of k-means clustering algorithm to predict student's learning activities. The information generated after the implementation of data mining technique may be helpful for instructor as well as for students.

Bray [12], in his study on private tutoring and its implications, observed that the percentage of students receiving private tutoring in India was relatively higher than in Malaysia, Singapore, Japan, China and Sri Lanka. It was also observed that there was an enhancement of academic performance with the intensity of private tutoring and this variation of intensity of private tutoring depends on the collective factor namely socio-economic conditions.

Bhardwaj and Pal [13] conducted study on the student performance based by selecting 300 students from 5 different degree college conducting BCA (Bachelor of Computer Application) course of Dr. R. M. L. Awadh University, Faizabad, India. By means of Bayesian classification method on 17 attribute, it was found that the factors like students' grade in senior secondary exam, living location, medium of teaching, mother's qualification, students other habit, family annual income and student's family status were highly correlated with the student academic performance.

IV. Data Mining Process

In present day's educational system, a students' performance is determined by the internal assessment and end semester examination. The internal assessment is carried out by the teacher based upon students' performance in educational activities such as class test, seminar, assignments, general proficiency, attendance and lab work. The end semester examination is one that is scored by the student in semester examination. Each student has to get minimum marks to pass a semester in internal as well as end semester examination.

A. Data Preparations

The data set used in this study was obtained from VBS Purvanchal University, Jaunpur (Uttar Pradesh) on the sampling method of computer Applications department of course MCA (Master of Computer Applications) from session 2007 to 2010. Initially size of the data is 50. In this step data stored in different tables was joined in a single table after joining process errors were removed.

B. Data selection and transformation

In this step only those fields were selected which were required for data mining. A few derived variables were selected. While some of the information for the variables was extracted from the database. All the predictor and response variables which were derived from the database are given in Table I for reference.

TABLE I. STUDENT RELATED VARIABLES

Variable	Description	Possible Values
PSM	Previous Semester Marks	{First > 60% Second >45 & <60% Third >36 & <45% Fail < 36%}
CTG	Class Test Grade	{Poor , Average, Good}
SEM	Seminar Performance	{Poor , Average, Good}
ASS	Assignment	{Yes, No}
GP	General Proficiency	{Yes, No}
ATT	Attendance	{Poor , Average, Good}
LW	Lab Work	{Yes, No}
ESM	End Semester Marks	{First > 60% Second >45 & <60% Third >36 & <45% Fail < 36%}

The domain values for some of the variables were defined for the present investigation as follows:

- **PSM –** Previous Semester Marks/Grade obtained in MCA course. It is split into five class values: *First – >60%, Second – >45% and <60%, Third – >36% and < 45%, Fail < 40%*.
- **CTG –** Class test grade obtained. Here in each semester two class tests are conducted and average of two class test are used to calculate sessional marks. CTG is split into three classes: *Poor – < 40%, Average – > 40% and < 60%, Good –>60%.*
- **SEM –** Seminar Performance obtained. In each semester seminar are organized to check the performance of students. Seminar performance is evaluated into three classes: *Poor – Presentation and communication skill is low, Average – Either presentation is fine or Communication skill is fine, Good – Both presentation and Communication skill is fine.*
- **ASS –** Assignment performance. In each semester two assignments are given to students by each teacher. Assignment performance is divided into two classes: *Yes – student submitted assignment, No – Student not submitted assignment.*
- **GP -** General Proficiency performance. Like seminar, in each semester general proficiency tests are organized. General Proficiency test is divided into two classes: *Yes – student participated in general proficiency, No – Student not participated in general proficiency.*
- **ATT –** Attendance of Student. Minimum 70% attendance is compulsory to participate in End Semester Examination. But even through in special cases low attendance students also participate in End Semester Examination on genuine reason. Attendance is divided into three classes: *Poor - <60%, Average - > 60% and <80%, Good - >80%.*
- **LW –** Lab Work. Lab work is divided into two classes: *Yes – student completed lab work, No – student not completed lab work.*
- **ESM -** End semester Marks obtained in MCA semester and it is declared as response variable. It is split into five class values: First – *>60% , Second – >45% and <60%, Third – >36% and < 45%, Fail < 40%.*

C. Decision Tree

A decision tree is a tree in which each branch node represents a choice between a number of alternatives, and each leaf node represents a decision.

Decision tree are commonly used for gaining information for the purpose of decision -making. Decision tree starts with a root node on which it is for users to take actions. From this node, users split each node recursively according to decision tree learning algorithm. The final result is a decision tree in which each branch represents a possible scenario of decision and its outcome.

The three widely used decision tree learning algorithms are: ID3, ASSISTANT and C4.5.

D. The ID3 Decision Tree

ID3 is a simple decision tree learning algorithm developed by Ross Quinlan [14]. The basic idea of ID3 algorithm is to construct the decision tree by employing a top-down, greedy search through the given sets to test each attribute at every tree node. In order to select the attribute that is most useful for classifying a given sets, we introduce a metric - information gain.

To find an optimal way to classify a learning set, what we need to do is to minimize the questions asked (i.e. minimizing the depth of the tree). Thus, we need some function which can measure which questions provide the most balanced splitting. The information gain metric is such a function.

E. Measuring Impurity

Given a data table that contains attributes and class of the attributes, we can measure homogeneity (or heterogeneity) of the table based on the classes. We say a table is pure or homogenous if it contains only a single class. If a data table contains several classes, then we say that the table is impure or heterogeneous. There are several indices to measure degree of impurity quantitatively. Most well known indices to measure degree of impurity are entropy, gini index, and classification error.

$$\text{Entropy} = \sum_j - p_j \log_2 p_j$$

Entropy of a pure table (consist of single class) is zero because the probability is 1 and log (1) = 0. Entropy reaches maximum value when all classes in the table have equal probability.

Gini Index = $1 - \sum_j p_j^2$

Gini index of a pure table consist of single class is zero because the probability is 1 and $1\text{-}1^2 = 0$. Similar to Entropy, Gini index also reaches maximum value when all classes in the table have equal probability.

Classification Error $= 1 - \max\{p_j\}$

Similar to Entropy and Gini Index, Classification error index of a pure table (consist of single class) is zero because the probability is 1 and 1-max (1) = 0. The value of classification error index is always between 0 and 1. In fact the maximum Gini index for a given number of classes is always equal to the maximum of classification error index because for a number of classes n, we set probability is equal to $p = \frac{1}{n}$ and maximum Gini index happens at $1 - n\frac{1}{n^2} = 1 - \frac{1}{n}$, while maximum classification error index also happens at $1 - \max\left\{\frac{1}{n}\right\} = 1 - \frac{1}{n}$.

F. Splitting Criteria

To determine the best attribute for a particular node in the tree we use the measure called Information Gain. The information gain, Gain (S, A) of an attribute A, relative to a collection of examples S, is defined as

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

Where *Values (A)* is the set of all possible values for attribute *A*, and S_v is the subset of *S* for which attribute *A* has value *v* (i.e., $S_v = \{s \in S \mid A(s) = v\}$). The first term in the equation for *Gain* is just the entropy of the original collection *S* and the second term is the expected value of the entropy after S is partitioned using attribute *A*. The expected entropy described by this second term is simply the sum of the entropies of each subset , weighted by the fraction of examples $\frac{|S_v|}{|S|}$ that belong to *Gain (S, A)* is therefore the expected reduction in entropy caused by knowing the value of attribute *A*.

$$\textit{Split Information (S, A)} = -\sum_{i=1}^{n} \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|}$$

and

$$\textit{Gain Ratio(S, A)} = \frac{Gain(S, A)}{Split\ Information(S, A)}$$

The process of selecting a new attribute and partitioning the training examples is now repeated for each non terminal descendant node. Attributes that have been incorporated higher in the tree are excluded, so that any given attribute can appear at most once along any path through the tree. This process continues for each new leaf node until either of two conditions is met:

1. Every attribute has already been included along this path through the tree, or
2. The training examples associated with this leaf node all have the same target attribute value (i.e., their entropy is zero).

G. The ID3Algoritm

ID3 (Examples, Target_Attribute, Attributes)

- Create a root node for the tree
- If all examples are positive, Return the single-node tree Root, with label = +.
- If all examples are negative, Return the single-node tree Root, with label = -.
- If number of predicting attributes is empty, then Return the single node tree Root, with label = most common value of the target attribute in the examples.
- Otherwise Begin
    - A = The Attribute that best classifies examples.
    - Decision Tree attribute for Root = A.
    - For each possible value, v_i, of A,
        - Add a new tree branch below Root, corresponding to the test A = v_i.
        - Let Examples(v_i) be the subset of examples that have the value v_i for A
        - If Examples(v_i) is empty
            - Then below this new branch add a leaf node with label = most common target value in the examples
        - Else below this new branch add the subtree ID3 (Examples(v_i), Target_Attribute, Attributes – {A})
- End
- Return Root

V. Results and Discussion

The data set of 50 students used in this study was obtained from VBS Purvanchal University, Jaunpur (Uttar Pradesh) Computer Applications department of course MCA (Master of Computer Applications) from session 2007 to 2010.

TABLE II. DATA SET

S. No.	PSM	CTG	SEM	ASS	GP	ATT	LW	ESM
1.	First	Good	Good	Yes	Yes	Good	Yes	First
2.	First	Good	Average	Yes	No	Good	Yes	First
3.	First	Good	Average	No	No	Average	No	First
4.	First	Average	Good	No	No	Good	Yes	First
5.	First	Average	Average	No	Yes	Good	Yes	First
6.	First	Poor	Average	No	No	Average	Yes	First

7.	First	Poor	Average	No	No	Poor	Yes	Second
8.	First	Average	Poor	Yes	Yes	Average	No	First
9.	First	Poor	Poor	No	No	Poor	No	Third
10.	First	Average	Average	Yes	Yes	Good	No	First
11.	Second	Good	Good	Yes	Yes	Good	Yes	First
12.	Second	Good	Average	Yes	Yes	Good	Yes	First
13.	Second	Good	Average	Yes	No	Good	No	First
14.	Second	Average	Good	Yes	Yes	Good	No	First
15.	Second	Good	Average	Yes	Yes	Average	Yes	First
16.	Second	Good	Average	Yes	Yes	Poor	Yes	Second
17.	Second	Average	Average	Yes	Yes	Good	Yes	Second
18.	Second	Average	Average	Yes	Yes	Poor	Yes	Second
19.	Second	Poor	Average	No	Yes	Good	Yes	Second
20.	Second	Average	Poor	Yes	No	Average	Yes	Second
21.	Second	Poor	Average	No	Yes	Poor	No	Third
22.	Second	Poor	Poor	Yes	Yes	Average	Yes	Third
23.	Second	Poor	Poor	No	No	Average	Yes	Third
24.	Second	Poor	Poor	Yes	Yes	Good	Yes	Second
25.	Second	Poor	Poor	Yes	Yes	Poor	Yes	Third
26.	Second	Poor	Poor	No	No	Poor	Yes	Fail
27.	Third	Good	Good	Yes	Yes	Good	Yes	First
28.	Third	Average	Good	Yes	Yes	Good	Yes	Second
29.	Third	Good	Average	Yes	Yes	Good	Yes	Second
30.	Third	Good	Good	Yes	Yes	Average	Yes	Second
31.	Third	Good	Good	No	No	Good	Yes	Second
32.	Third	Average	Average	Yes	Yes	Good	Yes	Second
33.	Third	Average	Average	No	Yes	Average	Yes	Third
34.	Third	Average	Good	No	No	Good	Yes	Third
35.	Third	Good	Average	No	Yes	Average	Yes	Third
36.	Third	Average	Poor	No	No	Average	Yes	Third
37.	Third	Poor	Average	Yes	No	Average	Yes	Third
38.	Third	Poor	Average	No	Yes	Poor	Yes	Fail
39.	Third	Average	Average	No	Yes	Poor	Yes	Third
40.	Third	Poor	Poor	No	No	Good	No	Third
41.	Third	Poor	Poor	No	Yes	Poor	Yes	Fail
42.	Third	Poor	Poor	No	No	Poor	No	Fail
43.	Fail	Good	Good	Yes	Yes	Good	Yes	Second
44.	Fail	Good	Good	Yes	Yes	Average	Yes	Second
45.	Fail	Average	Good	Yes	Yes	Average	Yes	Third
46.	Fail	Poor	Poor	Yes	Yes	Average	No	Fail
47.	Fail	Good	Poor	No	Yes	Poor	Yes	Fail
48.	Fail	Poor	Poor	No	No	Poor	Yes	Fail
49.	Fail	Average	Average	Yes	Yes	Good	Yes	Second
50.	Fail	Poor	Good	No	No	Poor	No	Fail

To work out the information gain for A relative to S, we first need to calculate the entropy of S. Here S is a set of 50 examples are 14 "*First*", 15 "*Second*", 13 "*Third*" and 8 "*Fail*"..

$$\text{Entropy (S)} = -p_{First}\log_2(p_{First}) - p_{Second}\log_2(p_{Second}) - p_{third}\log_2(p_{third}) - p_{Fail}\log_2(p_{Fail})$$

$$= -\left(\frac{14}{50}\right)\log_2\left(\frac{14}{50}\right) - \left(\frac{15}{50}\right)\log_2\left(\frac{15}{50}\right) - \left(\frac{13}{50}\right)\log_2\left(\frac{13}{50}\right) - \left(\frac{8}{50}\right)\log_2\left(\frac{8}{50}\right)$$

$$= 1.964$$

To determine the best attribute for a particular node in the tree we use the measure called Information Gain. The information gain, Gain (S, A) of an attribute A, relative to a collection of examples S,

$$Gain(S, PSM) = Entropy(S) - \frac{|S_{First}|}{|S|}Entropy(S_{First}) - \frac{|S_{Second}|}{|S|}Entropy(S_{Second}) - \frac{|S_{Third}|}{|S|}Entropy(S_{Third}) - \frac{|S_{Fail}|}{|S|}Entropy(S_{Fail})$$

TABLE III. GAIN VALUES

Gain	Value
Gain(S, PSM)	0.577036
Gain(S, CTG)	0.515173
Gain(S, SEM)	0.365881
Gain(S, ASS)	0.218628
Gain (S, GP)	0.043936
Gain(S, ATT)	0.451942
Gain(S, LW)	0.453513

PSM has the highest gain, therefore it is used as the root node as shown in figure 2.

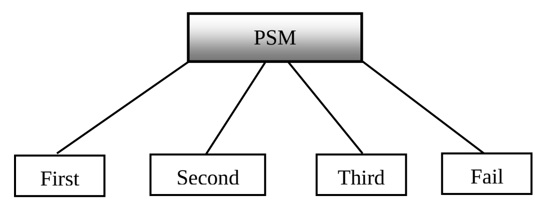


Figure 2. PSM as root node

Gain Ratio can be used for attribute selection, before calculating Gain ratio Split Information is shown in table IV.

TABLE IV. SPLIT INFORMATION

Split Information	Value
Split(S, PSM)	1.386579
Split (S, CTG)	1.448442
Split (S, SEM)	1.597734
Split (S, ASS)	1.744987
Split (S, GP)	1.91968
Split (S, ATT)	1.511673
Split (S, LW)	1.510102

Gain Ratio is shown in table V.

TABLE V. GAIN RATIO

Gain Ratio	Value
Gain Ratio (S, PSM)	0.416158
Gain Ratio (S, CTG)	0.355674
Gain Ratio (S, SEM)	0.229
Gain Ratio (S, ASS)	0.125289
Gain Ratio (S, GP)	0.022887
Gain Ratio (S, ATT)	0.298968
Gain Ratio (S, LW)	0.30032

This process goes on until all data classified perfectly or run out of attributes. The knowledge represented by decision tree can be extracted and represented in the form of IF-THEN rules.

IF PSM = ‘First’ AND ATT = ‘Good’ AND CTG = ‘Good’ or ‘Average’ THEN ESM = First
IF PSM = ‘First’ AND CTG = ‘Good’ AND ATT = “Good’ OR ‘Average’ THEN ESM = ‘First’
IF PSM = ‘Second’ AND ATT = ‘Good’ AND ASS = ‘Yes’ THEN ESM = ‘First’
IF PSM = ‘Second’ AND CTG = ‘Average’ AND LW = ‘Yes’ THEN ESM = ‘Second’
IF PSM = ‘Third’ AND CTG = ‘Good’ OR ‘Average’ AND ATT = “Good’ OR ‘Average’ THEN PSM = ‘Second’
IF PSM = ‘Third’ AND ASS = ‘No’ AND ATT = ‘Average’ THEN PSM = ‘Third’
IF PSM = ‘Fail’ AND CTG = ‘Poor’ AND ATT = ‘Poor’ THEN PSM = ‘Fail’

Figure 3. Rule Set generated by Decision Tree

One classification rules can be generated for each path from each terminal node to root node. Pruning technique was executed by removing nodes with less than desired number of objects. IF- THEN rules may be easier to understand is shown in figure 3.

CONCLUSION

In this paper, the classification task is used on student database to predict the students division on the basis of previous database. As there are many approaches that are used for data classification, the decision tree method is used here. Information’s like Attendance, Class test, Seminar and Assignment marks were collected from the student’s previous database, to predict the performance at the end of the semester.

This study will help to the students and the teachers to improve the division of the student. This study will also work to identify those students which needed special attention to reduce fail ration and taking appropriate action for the next semester examination.

REFERENCES

[1] Heikki, Mannila, Data mining: machine learning, statistics, and databases, IEEE, 1996.

[2] U. Fayadd, Piatesky, G. Shapiro, and P. Smyth, From data mining to knowledge discovery in databases, AAAI Press / The MIT Press, Massachusetts Institute Of Technology. ISBN 0–262 56097–6, 1996.

[3] J. Han and M. Kamber, “Data Mining: Concepts and Techniques,” Morgan Kaufmann, 2000.

[4] Alaa el-Halees, “Mining students data to analyze e-Learning behavior: A Case Study”, 2009..

[5] U . K. Pandey, and S. Pal, “Data Mining: A prediction of performer or underperformer using classification”, (IJCSIT) International Journal of Computer Science and Information Technology, Vol. 2(2), pp.686-690, ISSN:0975-9646, 2011.

[6] S. T. Hijazi, and R. S. M. M. Naqvi, “Factors affecting student’s performance: A Case of Private Colleges”, Bangladesh e-Journal of Sociology, Vol. 3, No. 1, 2006.

[7] Z. N. Khan, “Scholastic achievement of higher secondary students in science stream”, Journal of Social Sciences, Vol. 1, No. 2, pp. 84-87, 2005..

[8] Galit.et.al, “Examining online learning processes based on log files analysis: a case study”. Research, Reflection and Innovations in Integrating ICT in Education 2007.

[9] Q. A. AI-Radaideh, E. W. AI-Shawakfa, and M. I. AI-Najjar, “Mining student data using decision trees”, International Arab Conference on Information Technology(ACIT'2006), Yarmouk University, Jordan, 2006.

[10] U. K. Pandey, and S. Pal, “A Data mining view on class room teaching language”, (IJCSI) International Journal of Computer Science Issue, Vol. 8, Issue 2, pp. 277-282, ISSN:1694-0814, 2011.

[11] Shaeela Ayesha, Tasleem Mustafa, Ahsan Raza Sattar, M. Inayat Khan, “Data mining model for higher education system”, Europen Journal of Scientific Research, Vol.43, No.1, pp.24-29, 2010.

[12] M. Bray, The shadow education system: private tutoring and its implications for planners, (2nd ed.), UNESCO, PARIS, France, 2007.

[13] B.K. Bharadwaj and S. Pal. “Data Mining: A prediction for performance improvement using classification”, International Journal of Computer Science and Information Security (IJCSIS), Vol. 9, No. 4, pp. 136-140, 2011.

[14] J. R. Quinlan, “Introduction of decision tree: Machine learn”, 1: pp. 86-106, 1986.

[15] Vashishta, S. (2011). Efficient Retrieval of Text for Biomedical Domain using Data Mining Algorithm. *IJACSA - International Journal of Advanced Computer Science and Applications*, *2*(4), 77-80.

[16] Kumar, V. (2011). An Empirical Study of the Applications of Data Mining Techniques in Higher Education. *IJACSA - International Journal of Advanced Computer Science and Applications*, *2*(3), 80-84. Retrieved from http://ijacsa.thesai.org.

AUTHORS PROFILE

Brijesh Kumar Bhardwaj is Assistant Professor in the Department of Computer Applications, Dr. R. M. L. Avadh University Faizabad India. He obtained his M.C.A degree from Dr. R. M. L. Avadh University Faizabad (2003) and M.Phil. in Computer Applications from Vinayaka mission University, Tamilnadu. He is currently doing research in Data Mining and Knowledge Discovery. He has published one international paper.

Saurabh Pal received his M.Sc. (Computer Science) from Allahabad University, UP, India (1996) and obtained his Ph.D. degree from the Dr. R. M. L. Awadh University, Faizabad (2002). He then joined the Dept. of Computer Applications, VBS Purvanchal University, Jaunpur as Lecturer. At present, he is working as Head and Sr. Lecturer at Department of Computer Applications.

Saurabh Pal has authored a commendable number of research papers in international/national Conference/journals and also guides research scholars in Computer Science/Applications. He is an active member of IACSIT, CSI, Society of Statistics and Computer Applications and working as Reviewer/Editorial Board Member for more than 15 international journals. His research interests include Image Processing, Data Mining, Grid Computing and Artificial Intelligence.

Comparison between Traditional Approach and Object-Oriented Approach in Software Engineering Development

Nabil Mohammed Ali Munassar 1
PhD Student 3rd year of Computer Science & Engineering
Jawaharlal Nehru Technological University
Kuktapally, Hyderabad- 500 085, Andhra Pradesh, India

Dr. A. Govardhan 2
Professor of Computer Science & Engineering
Principal JNTUH of Engineering College, Jagityal,
Karimnagar (Dt), A.P., India

***Abstract*— This paper discusses the comparison between Traditional approaches and Object-Oriented approach.**

Traditional approach has a lot of models that deal with different types of projects such as waterfall, spiral, iterative and v-shaped, but all of them and other lack flexibility to deal with other kinds of projects like Object-Oriented.

Object–oriented Software Engineering (OOSE) is an object modeling language and methodology. The approach of using object – oriented techniques for designing a system is referred to as object–oriented design. Object–oriented development approaches are best suited to projects that will imply systems using emerging object technologies to construct, manage, and assemble those objects into useful computer applications. Object oriented design is the continuation of object-oriented analysis, continuing to center the development focus on object modeling techniques.

Keywordss- Software Engineering; Traditional Approach; Object-Oriented Approach; Analysis; Design; Deployment; Test; methodology; Comparison between Traditional Approach and Object-Oriented Approach.

I. Introduction

All software, especially large pieces of software produced by many people, should be produced using some kind of methodology. Even small pieces of software developed by one person can be improved by keeping a methodology in mind. A methodology is a systematic way of doing things. It is a repeatable process that we can follow from the earliest stages of software development through to the maintenance of an installed system. As well as the process, a methodology should specify what we're expected to produce as we follow the process. A methodology will also include recommendation or techniques for resource management, planning, scheduling and other management tasks. Good, widely available methodologies are essential for a mature software industry.

A good methodology will address at least the following issues: Planning, Scheduling, Resourcing, Workflows, Activities, Roles, Artifacts, Education. There are a number of phases common to every development, regardless of methodology, starting with requirements capture and ending with maintenance. During the last few decades a number of software development models have been proposed and discussed within the Software Engineering community. With the traditional approach, you're expected to move forward gracefully from one phase to the other. With the modern approach, on the other hand, you're allowed to perform each phase more than once and in any order [1, 10].

II. Traditional Approach

There are a number of phases common to every development, regardless of methodology, starting with requirements capture and ending with maintenance. With the traditional approach, will be expected to move forward gracefully from one phase to the other. The list below describes the common phases in software development [1, 6].

A. Requirements

Requirements capture is about discovering what is going to achieve with new piece of software and has two aspects. Business modeling involves understanding the context in which software will operate. A system requirement modeling (or functional specification) means deciding what capabilities the new software will have and writing down those capabilities [1].

B. Analysis

Analysis means understanding what are dealing with. Before designing a solution, it needs to be clear about the relevant entities, their properties and their inter-relationships. Also needs to be able to verify understanding. This can involve customers and end users, since they're likely to be subject-matter experts [1].

C. Design

In the design phase, will work out, how to solve the problem. In other words, make decisions based on experience, estimation and intuition, about what software which will write and how will deploy it. System design breaks the system down into logical subsystems (processes) and physical subsystems (computers and networks), decides how machines will communicate, and chooses the right technologies for the job, and so on [1].

D. Specification

Specification is an often-ignored, or at least often-neglected, phase. The term specification is used in different ways by different developers. For example, the output of the requirements phase is a specification of what the system must be able to do; the output of analysis is a specification of what are dealing with; and so on [3].

E. Implementation

In this phase is writing pieces of code that work together to form subsystems, which in turn collaborate to form the whole system. The sort of the task which is carried out during the implementation phase is 'Write the method bodies for the Inventory class, in such a way that they conform to their specification' [5].

F. Testing

When the software is complete, it must be tested against the system requirements to see if it fits the original goals. It is a good idea for programmers to perform small tests as they go along, to improve the quality of the code that they deliver [5].

G. Deployment

In the deployment phase, are concerned with getting the hardware and software to the end users, along with manuals and training materials. This may be a complex process, involving a gradual, planned transition from the old way of working to the new one [1].

H. Maintenance

When the system is deployed, it has only just been born. A long life stretches before it, during which it has to stand up to everyday use – this is where the real testing happens. The sort of the problem which is discovered discover during the maintenance phase is 'When the log-on window opens, it still contains the last password entered.' As the software developers, we normally interested in maintenance because of the faults (bugs) that are found in software. Must find the faults and remove them as quickly as possible, rolling out fixed versions of the software to keep the end users happy. As well as faults, users may discover deficiencies (things that the system should do but doesn't) and extra requirements (things that would improve the system) [3, 6].

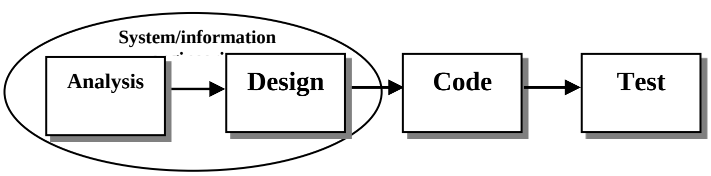


Figure 1: The linear Sequential Model [6].

III. OBJECT-ORIENTED APPROACH

In object-oriented approach, a system is viewed as a set of objects. All object-orientation experts agree that a good methodology is essential for software development, especially when working in teams. Thus, quite a few methodologies have been invented over the last decade. Broadly speaking, all object-oriented methodologies are alike – they have similar phases and similar artifacts – but there are many small differences. Object-oriented methodologies tend not to be too prescriptive: the developers are given some choice about whether they use a particular type of diagram, for example. Therefore, the development team must select a methodology and agree which artifacts are to be produced, before they do any detailed planning or scheduling. In general, each methodology addresses:

- The philosophy behind each of the phases.
- The workflows and the individual activities within each phase.
- The artifacts that should be produced (diagrams, textual descriptions and code).
- Dependencies between the artifacts.
- Notations for the different kinds of artifacts.
- The need to model static structure and dynamic behavior.

Static modeling involves deciding what the logical or physical parts of the system should be and how they should be connected together. Dynamic modeling is about deciding how the static parts should collaborate. Roughly speaking, static modeling describes how we construct and initialize the system, while dynamic modeling describes how the system should behave when it's running. Typically, we produce at least one static model and one dynamic model during each phase of the development.

Some methodologies, especially the more comprehensive ones, have alternative development paths, geared to different types and sizes of development.[1,4]

The benefits of Object-Oriented Development are reduced time to market, greater product flexibility, and schedule predictability and the risks of them are performance and start-up costs [5].

A. Analysis

The aim of the analysis process is to analyze, specify, and define the system which is to be built. In this phase, we build models that will make it easier for us to understand the system. The models that are developed during analysis are oriented fully to the application and not the implementation environment; they are "essential" models that are independent of such things as operating system, programming language, DBMS, processor distribution, or hardware configuration.

Two different models are developed in analysis; the Requirements Model and the Analysis Model. These are based on requirement specifications and discussions with the prospective users. The first model, the Requirements Model, should make it possible to delimit the system and to define what functionality should take place within it. For this purpose we develop a conceptual picture of the system using problem domain objects and also specific interface descriptions of the system if it is meaningful for this system. We also describe the system as a number of use cases that are performed by a number of actors. The Analysis Model is an architectural model used for analysis of robustness. It gives a conceptual configuration of the system, consisting of various object classes: active controllers, domain entities, and interface objects. The purpose of this model is to find a robust and extensible structure for the system as a base for construction. Each of the object types has its own special purpose for this robustness, and together they will offer the total functionality that was specified in the Requirements Model. To manage the development, the Analysis Model may combine objects into Subsystems [2].

B. Construction

We build our system through construction based on the Analysis Model and the Requirements Model created by the analysis process. The construction process lasts until the coding is completed and the included units have been tested. There are three main reasons for a construction process:

1) *The Analysis Model is not sufficiently formal.*
2) *Adaptation must be made to the actual implementation environment.*
3) *We want to do internal validation of the analysis results.*

The construction activity produces two models, the Design Model and the Implementation Model. Construction is thus divided into two phases; design and implementation, each of which develops a model. The Design Model is a further refinement and formalization of the Analysis Model where consequences of the implementation environment have been taken into account. The Implementation model is the actual implementation (code) of the system. [2].

C. Testing

Testing is an activity to verify that a correct system is being built. Testing is traditionally an expensive activity, primarily because many faults are not detected until late in the development. To do effective testing we must have as a goal that every test should detect a fault.

Unit testing is performed to test a specific unit, where a unit can be of varying size from a class up to an entire subsystem. The unit is initially tested structurally, that is, "white box testing." This means that we use our knowledge of the inside of the unit to test it. We have various coverage criteria for the test, the minimum being to cover all statements. However, coverage criteria can be hard to define, due to polymorphism; many branches are made implicit in an object-oriented system. However, polymorphism also enhances the independence of each object, making them easier to test as standalone units. The use of inheritance also complicates testing, since we may need to retest operations at different levels in the inheritance hierarchy. On the other hand, since we typically have less code, there is less to test. Specification testing of a unit is done primarily from the object protocol (so-called "black box testing). Here we use equivalence partitioning to find appropriate test cases. Test planning must be done early, along with the identification and specification of tests [2].

D. UML

By the mid-1990s, the best-known methodologies were those invented by Ivar Jacobson, James Rumbaugh and Grady Booch. Each had his own consulting company using his own methodology and his own notation. By 1996, Jacobson and Rumbaugh had joined Rational Corporation, and they had developed a set of notations which became known as the Unified Modeling Language (UML). The ‘three amigos’, as they have become known, donated UML to the Object Management Group (OMG) for safekeeping and improvement. OMG is a not-for-profit industry consortium, founded in 1989 to promote open standards for enterprise-level object technology; their other well-known work is CORBA [1].

1) Use Case Diagram

A use case is a static description of some way in which a system or a business is used, by its customers, its users or by other systems. A use case diagram shows how system use cases are related to each other and how the users can get at them. Each bubble on a use case diagram represents a use case and each stick person represents a user. Figure 2 depicts a car rental store accessible over the Internet. From this picture, we can extract a lot of information quite easily. For example, an Assistant can make a reservation; a Customer can look for car models; Members can log on; users must be logged on before they can make reservations; and so on [1, 3].

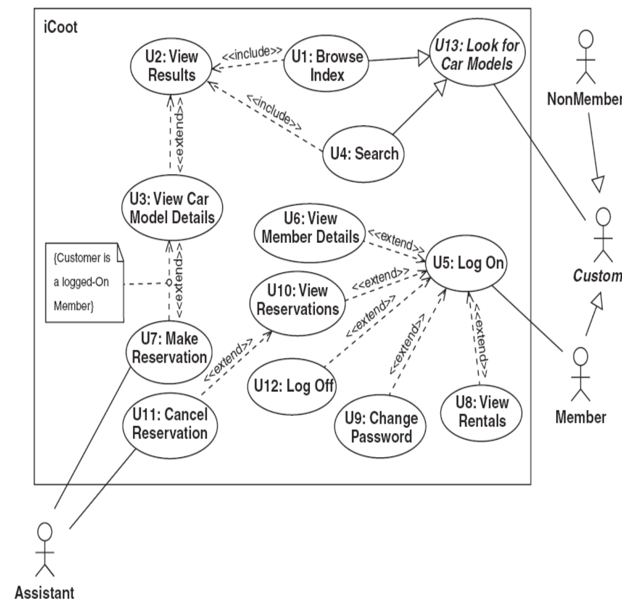


Figure 2: A use Case Diagram

2) Class Diagram (Analysis Level)

A class diagram shows which classes exist in the business (during analysis) or in the system itself (during subsystem design). Figure 3 shows an example of an analysis-level class diagram, with each class represented as a labeled box. As well as the classes themselves, a class diagram shows how objects of these classes can be connected together. For example, Figure 3 shows that a CarModel has inside it a CarModelDetails, referred to as its details.U3: View Car Model Details. (Extends U2, extended by U7.) Preconditions: None.

a) *Customer selects one of the matching Car Models.*
b) *Customer requests details of the selected Car Model.*
c) *iCoot displays details of the selected Car Model (make, engine size, price, description, advert and poster).*
d) *If Customer is a logged-on Member, extend with U7.*

Postconditions: iCoot has displayed details of selected Car Models.

Nonfunctional Requirements: r1. Adverts should be displayed using a streaming protocol rather than requiring a download [1, 5].

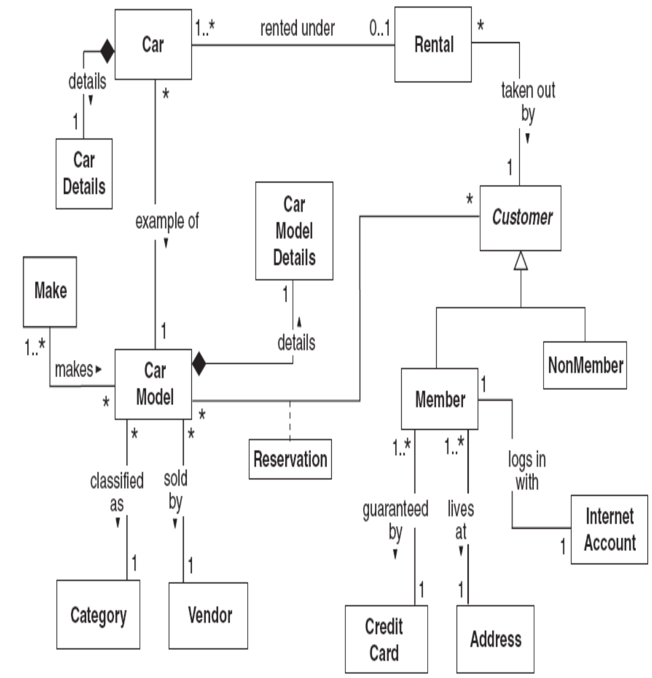


Figure 3: A class Diagram at the Analysis Level.

3) Communication Diagram

A communication diagram, as its name suggests, shows collaborations between objects. The one shown in Figure 4 describes the process of reserving a car model over the Internet: A Member tells the MemberUI to reserve a CarModel; the MemberUI tells the ReservationHome to create a Reservation for the given CarModel and the current Member; the MemberUI then asks the new Reservation for its number and returns this to the Member [1].

1) Deployment Diagram

A deployment diagram shows how the completed system will be deployed on one or more machines. A deployment diagram can include all sorts of features such as machines, processes, files and dependencies. Figure 5 shows that any number of HTMLClient nodes (each hosting a Web Browser) and GUIClient nodes communicate with two server machines, each hosting a WebServer and a CootBusinessServer; each Web Server communicates with a CootBusinessServer; and each CootBusinessServer communicates with a DBMS running on one of two DBServer nodes [1].

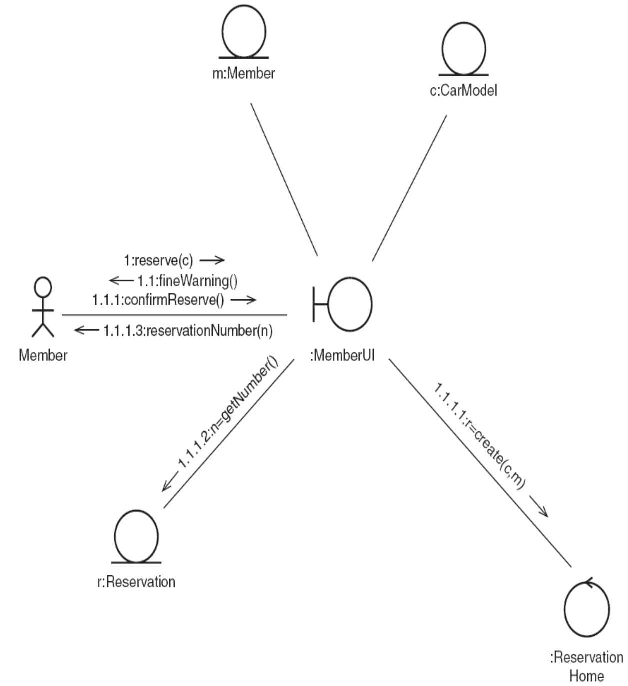


Figure 4: A communication Diagram

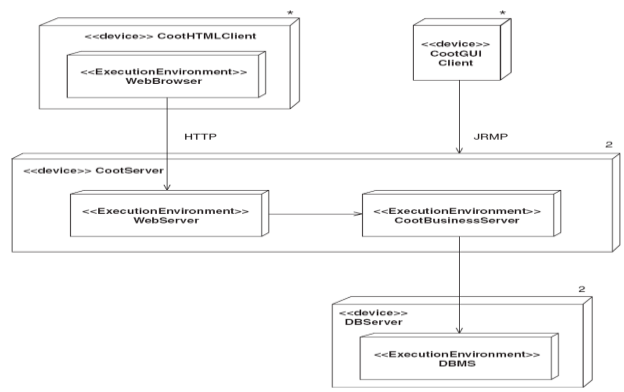


Figure 5: A deployment Diagram.

2) Class Diagram (Design Level)

The class diagram shown in Figure 6 uses the same notation as the one introduced in Figure 3. The only difference is that design-level class diagrams tend to use more of the available notation, because they are more detailed. This one expands on part of the analysis class diagram to show methods, constructors and navigability [1, 3].

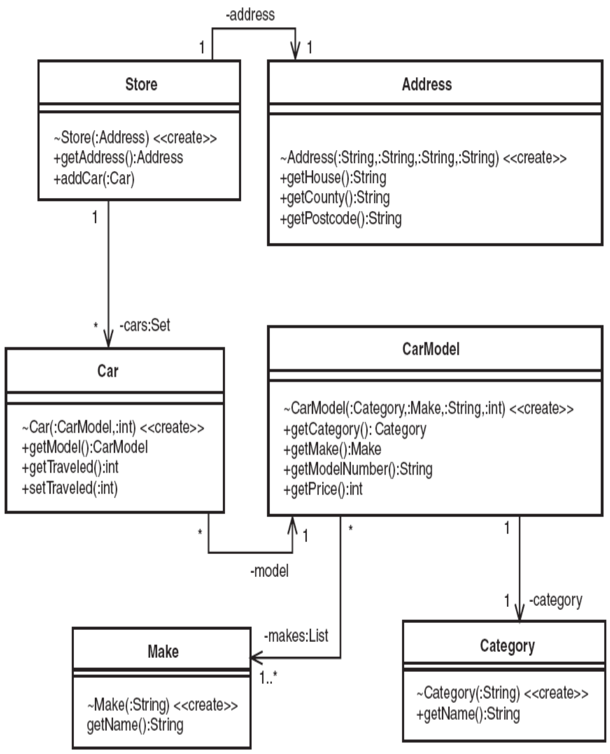


Figure 6: A design-level Class Diagram

3) *Sequence Diagram*

A sequence diagram shows interactions between objects. Communication diagrams also show interactions between objects, but in a way that emphasizes links rather than sequence. Sequence diagrams are used during subsystem design, but they are equally applicable to dynamic modeling during analysis, system design and even requirements capture. The diagram in Figure 7 specifies how a Member can log off from the system. Messages are shown as arrows flowing between vertical bars that represent objects (each object is named at the top of its bar). Time flows down the page on a sequence diagram. So, Figure 7 specifies, in brief: a Member asks the AuthenticationServlet to logoff; the AuthenticationServlet passes the request on to the AuthenticationServer, reading the id from the browser session; the AuthenticationServer finds the corresponding Member object and tells it to set its session id to 0; the Member passes this request on to its InternetAccount; finally, the Member is presented with the home page [1, 5].

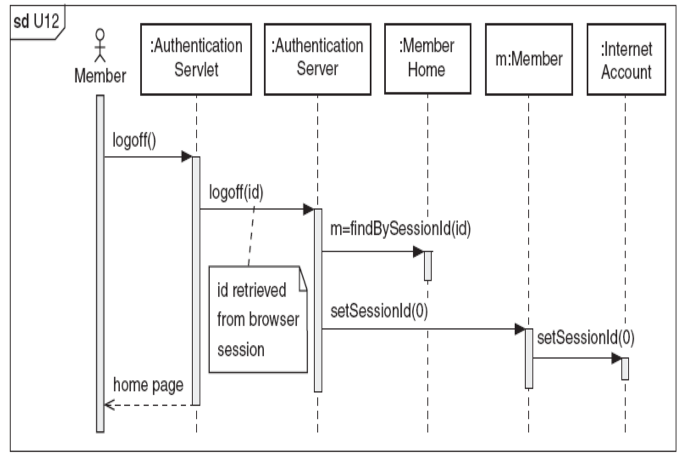


Figure 7: A sequence Diagram from the Design Phase

IV. COMPARISON BETWEEN TRADITONAL APPROACH AND OBJECT-ORIENTED APPROACH TO DEVELOPMENT IN SOFTWARE ENGINEERING

Summarize the comparison between Traditional Approach and Object-Oriented Approach shows through the table1.

TABLE 1. COMPARISON BETWEEN TRADITIONAL APPROACH AND OBJECT-ORIENTED APPROACH

TABLE I.

Traditional Approach	Object-Oriented Approach
Used to develop the Traditional Projects that uses procedural programming.	Used to develop Object-oriented Projects that depends on Object-Oriented programming.
Uses common processes likes: analysis, design, implementation, and testing.	Uses UML notations likes: use case, class diagram, communication diagram, development diagram and sequence diagram.
Depends on the size of projects and type of projects.	Depends on the experience of the team and complexity of projects through the numbers of objects.
Needs to large duration sometimes to development the large projects.	Need to more time than Traditional approach and leads that to more cost.
The problem of Traditional approach using classical life cycle [7, 8].	The object-oriented software life cycle identifies the three traditional activities of analysis, design, and implementation.[8].

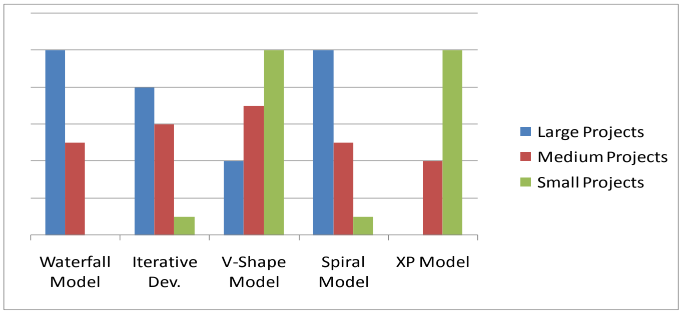


Figure 8: Illustrate The Different Models of Traditional Approach with Different Projects. [1, 6, 11]

From the previous figure 8 which illustrates the five models from traditional approach that deals with three types of projects, where we notice the waterfall model deals properly with large and medium projects like spiral model and iterative model that needs more time more cost and experience for team, however the V-shape model and XP model use properly with medium and small projects, because they need little time and some experience of team to perform projects.

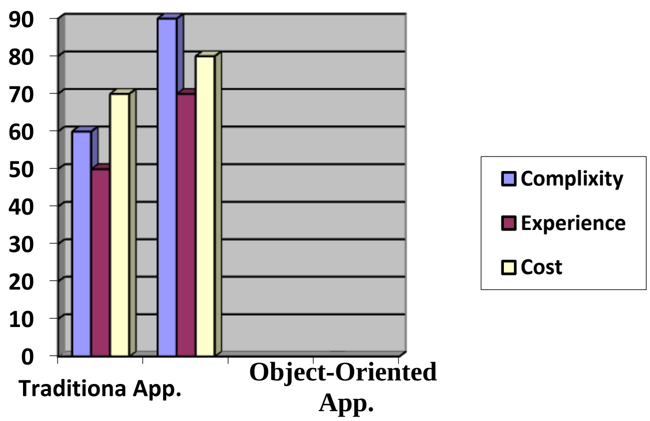


Figure 9: Illustrate The Different Criteria (Complexity, Experience and Cost) for Traditional Approach and Object-oriented Approach. [3, 5, 10]

From the previous chart illustrates the some criteria such as (Complexity, Experience, and Cost). In Traditional Approach this criterion depends on the type of model and size of project, but in general as shows from figure 9 is little above from the middle, however the Object-Oriented Approach depends on the complexity of project that leads to increase the cost than other approach.

V. Conclusion and Future Work

After completing this paper, it is concluded that:

1. As with any technology or tool invented by human beings, all SE methodologies have limitations [9].
2. The software engineering development has two ways to develop the projects that: traditional approach and object-oriented approach.
3. The traditional approach uses traditional projects that used in development of their procedural programming like C, this approach leads software developers to focus on Decomposition of larger algorithms into smaller ones.
4. The object-oriented approach uses to development the object-oriented projects that use the object-oriented programming like: C++ and Java.
5. The object-oriented approach to software development has a decided advantage over the traditional approach in dealing with complexity and the fact that most contemporary languages and tools are object-oriented.

Finally, some topics can be suggested for future works:

1. Design the model that includes the features of traditional approach and object-oriented approach to develop and deals with different projects in software engineering.
2. Updating some traditional approach to be able to use different types of projects.
3. Simplifying the object-oriented approach through its steps to use the smallest projects that deal with simple programming.

References

[1] Mike O'Docherty, "Object-Oriented Analysis and Design Understanding System Development with UML 2.0", John Wiley & Sons Ltd, England, 2005.
[2] Magnus Christerson and Larry L. Constantine, "Object-Oriented Software Engineering- A Use Case Driven Approach ", Objective Systems, Sweden, 2009.
[3] Ian sommerville, "Software Engineering", Addison Wesley, 7th edition, 2004.
[4] Pankaj Jalote, "An Integrated Approach to Software Engineering", Springer Science Business Media, Inc, Third Edition, 2005.
[5] Grady Booch, "Object-Oriented Analysis and Design with applications", Addison Wesley Longman, Inc, second Edition, 1998.
[6] Roger S. Pressman, "Software Engineering a practitioner's approach", McGraw-Hill, 5th edition, 2001.
[7] M M Lehman,"Process Models, Process Programs, Programming Support", ACM, 1987.
[8] Tim Korson and John D. McGregor," Understanding Object-Oriented: A Unifying Paradigm", ACM, Vol. 33, No. 9, 1990.
[9] Li Jiang and Armin Eberlein," Towards A Framework for Understanding the Relationships between Classical Software Engineering and Agile Methodologies", ACM, 2008.
[10] Luciano Rodrigues Guimarães and Dr. Plínio Roberto Souza Vilela," Comparing Software Development Models Using CDM", ACM, 2005.
[11] Alan M. Davis and Pradip Sitaram, "A Concurrent Process Model of Software Development", ACM, Software Engineering Notes Vol. 19 No. 2, 1994.

Authors Profile

Nabil Mohammed Ali Munassar

Was born in Jeddah, Saudi Arabia in 1978. He studied Computer Science at University of Science and Technology, Yemen from 1997 to 2001. In 2001 he received the Bachelor degree. He studied Master of Information Technology at Arab Academic, Yemen, from 2004 to 2007. Now he Ph.D. Student 3rd year of CSE at Jawaharlal Nehru Technological University (JNTU), Hyderabad, A. P., India. He is working as Associate Professor in Computer Science & Engineering College in University Of Science and Technology, Yemen. His areas of interest include Software Engineering, System Analysis and Design, Databases and Object Oriented Technologies.

Dr.A.Govardhan

Received Ph.D. degree in Computer Science and Engineering from Jawaharlal Nehru Technological University in 2003, M.Tech. from Jawaharlal Nehru University in 1994 and B.E. from Osmania University in 1992. He is working as a Principal of Jawaharlal Nehru Technological University, Jagitial. He has published around 108 papers in various national and international Journals/conferences. His research of interest includes Databases, Data Warehousing & Mining, Information Retrieval, Computer Networks, Image Processing, Software Engineering, Search Engines and Object Oriented Technologies.

Estimation of Dynamic Background and Object Detection in Noisy Visual Surveillance

M.Sankari
Department of Computer Applications,
Nehru Institute of Engineering and Technology,
Coimbatore, INDIA

C. Meena
Head, Computer Centre,
Avinashilingam University,
Coimbatore, INDIA

***Abstract*—Dynamic background subtraction in noisy environment for detecting object is a challenging process in computer vision. The proposed algorithm has been used to identify moving objects from the sequence of video frames which contains dynamically changing backgrounds in the noisy atmosphere. There are many challenges in achieving a robust background subtraction algorithm in the external noisy environment. In connection with our previous work, in this paper, we have proposed a methodology to perform background subtraction from moving vehicles in traffic video sequences that combines statistical assumptions of moving objects using the previous frames in the dynamically varying noisy situation. Background image is frequently updated in order to achieve reliability of the motion detection. For that, a binary moving objects hypothesis mask is constructed to classify any group of lattices as being from a moving object based on the optimal threshold. Then, the new incoming information is integrated into the current background image using a Kalman filter. In order to improve the performance, a post-processing has been done. It has been accomplished by shadow and noise removal algorithms operating at the lattice which identifies object-level elements. The results of post-processing can be used to detect object more efficiently. Experimental results and analysis show the prominence of the proposed approach which has achieved an average of 94% accuracy in real-time acquired images.**

Keywords- Background subtraction; Background updation; Binary segmentation mask; Kalman filter; Noise removal; Shadow removal; Traffic video sequences.

I. INTRODUCTION

In visual surveillance model, estimating the dynamic background and detecting the object from the noisy environment is a computationally challenging problem. Our main target is to identify the object from the multi model background using background subtraction, shadow removal and noise removal techniques. For that we need to detect and extract the foreground object from the background image. After detecting the foreground object there may a large number of possible degradations that an image can suffer. Common degradations are blurring, motion and noise. Blurring can be caused when an object in the image is outside the cameras due to loss of depth information during the exposure. In the proposed approach after detecting the object image, converting it into its spatial frequencies, developing a point spread function (PSF) to filter the image with, and then converting the filtered result back into the spatial domain to see if blur was removed.

This can be done in several steps. At the end, an algorithm was developed for removing blur from an already blurry image with no information regarding the blurring PSF. In-class variability, occlusion, and lighting conditions also change the overall appearance of vehicles. Region along the road changes continuously while the lighting conditions depend on the time of the day and the weather. The entire process is automatic and uses computation time that scales according to the size of the input video sequence.

The remainder of the paper is organized as follows: Section II gives the overview of the related work. Section III describes the architecture and modeling of proposed methodology for background elimination and object detection. Implementation and performance are analyzed in section IV. Section V contains the concluding remarks and future work.

II. OVERVIEW OF THE RELATED WORK

Scores of research have been done in the literature in order to attain a solution to an efficient and reliable background subtraction. To detect moving objects in a dynamic scene, adaptive background subtraction techniques have been developed [1] [2] [3]. Adaptive Gaussian mixtures are commonly chosen for their analytical representation and theoretical foundations. For these reasons, they have been employed in real-time surveillance systems for background subtraction [4] [5] and object tracking [6]. For foreground analysis [7] [8], a method for foreground analysis was proposed for moving object, shadow, and ghost by combining the motion information. The computation cost is relatively expensive for real-time video surveillance systems because of the computation of optical flow. In [9], a work has presented on a novel background subtraction algorithm that is capable of detecting objects of interest while all pixels are in motion. Background subtraction technique is mostly used for motion pictures to segment the foreground object by most of the researchers [10] [11]. Liyuan Li, et al. [12] proposed foreground object detection through foreground and background classifications under bayesian framework. In addition, moving object segmentation with background suppression is affected by the problem of shadows [6] [13]. Indeed, the moving object detection do not classify shadows as belonging to foreground objects since the appearance and geometrical properties of the object can be distorted which, in turn, affects many subsequent tasks such as object classification and the assessment of moving object position. In this paper, we propose a novel simple method that exploits all these features, combining them so as to efficiently provide detection of moving objects, ghosts, and shadows. The main contribution of this proposal is the integration of knowledge of detected objects, shadows, and ghosts in the segmentation process to enhance both object segmentation and background keep posted. The resulting method proves to be accurate and reactive and, at the same time, fast and flexible in the applications.

III. Proposed Work

The proposed system extracts foreground objects such as people, objects, or events of interest in variety of noisy environment. The schematic flow of the proposed algorithm is shown in Fig.1. This is an extension work of our previous method [14]. Typically, these systems consist of stationary cameras placed in the highways. These cameras are integrated with, intelligent computer systems that perform preprocessing operation from the captured video images and notify human operators or trigger control process. The objective of this real-time motion detection and tracking algorithm is to provide low-level functionality for building higher-level recognition capabilities.

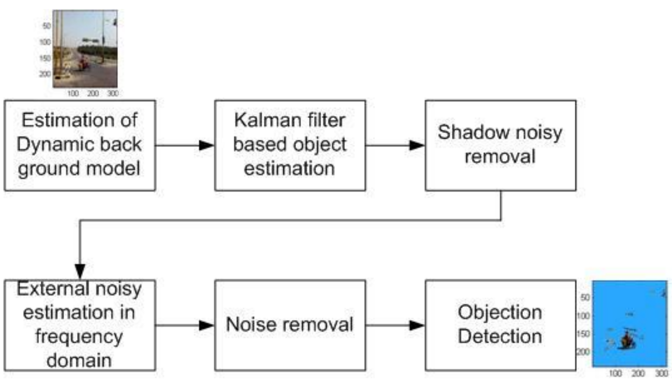


Figure 1. Schematic flow of the proposed algorithm to detect the objects in the noisy environment.

A. Preprocessing

Preprocessing is the key step and the starting point for image analysis, due to the wide diversity of resolution, image format, sampling models, and illumination techniques that are used during acquisition. In our method, preprocessing step was done by statistical method using adaptive median filter. The resultant frames are then utilized as an input for the background subtraction module. Image *I(x,y)* at time *t* is shown in Fig.2. The background image *B(x,y)* at time *t* is shown Fig.3.

Figure 2. Image at time *t*: *I (x; y; t).*

Figure 3. Image at time *t*: *B (x; y; t).*

In order to get the estimated background, we have used an adaptive median filter. Basically, impulse noise is a major artifact that affects the sequence of frame in the surveillance system. For this reason to estimate the background in the noisy environment we have proposed an adaptive median filter (AMF). The AMF can be used to enhance the quality of noisy signals, in order to achieve better forcefulness in pattern recognition and adaptive control systems. It executes on spatial processing to determine which pixels in an image have been exaggerated by impulse noise. The AMF categorizes pixels as noise by contrasting each pixel in the image to its close proximity of neighbor pixels. The size of the neighborhood and its threshold are adaptable for the robust assessment. A pixel that is dissimilar from a mainstream of its neighbors, as well as being not logically aligned with those pixels to which it is similar, is labeled as impulse noise. These noise pixels are then substituted by the median pixel value of the pixels in the neighborhood that have passed the noise labeling test [15]. The following steps were used for back ground estimation.

Step 1: Estimate the background at time t using adaptive median filter method.

Step 2: Subtract the estimated background from the input frame.

Step 3: Apply a threshold φ to the absolute difference to get the binary moving objects hypothesis mask.

Assuming that the background is more likely to appear in a scene, we can use the median of the previous n frames as the background model

$$B(x,y,t) = \Theta((I(x,y,t-i)), \quad (1)$$

$$|I(x,y,t) = \Theta(I(x,y,t-i))| > \varphi, \quad (2)$$

where $i \in \{0,1,2,...,n-1\}$ and computation of $\Theta(.)$ is based on the AMF. The following are the algorithm for median filter computation. Algorithm consists of two steps. Step 1 is described for deciding whether the median of the gray values in the size of the neighborhood or not.

Step 1: $$\begin{aligned} \varpi_1 &= \Theta(\delta(x,y)) - \delta_{\min}(x,y) \\ \varpi_2 &= \Theta(\delta(x,y)) - \delta_{\max}(x,y) \end{aligned},$$

If $\varpi_1 > 0$ && $\varpi_2 < 0$ then do the Step 2 otherwise enlarge the neighborhood window size based on the maximum allowed size of the input image. Step 1 is executed until $S_w > S_{\max}$ otherwise step 2 is computed.

Step 2: $\begin{array}{l} \varpi_1 = G(x,y) - \delta_{\min}(x,y) \\ \varpi_2 = G(x,y) - \delta_{\max}(x,y) \end{array}$,

If $\varpi_1 > 0$ && $\varpi_2 < 0$ then we have taken $G(x,y)$ as an output else $\Theta(\delta(x,y))$.

where $\delta(x,y)$ is the AMF neighborhood size. $\delta_{\min}(x,y)$ is the minimum gray level value in neighborhood $\delta(x,y)$, $\delta_{\max}(x,y)$ represents the maximum gray level value in $\delta(x,y)$, $\Theta(\delta(x,y))$ denotes the median of gray levels in $\delta(x,y)$, $G(x,y)$ is the gray level at coordinates (x,y), S_w represents window size of the current process and $S_{\max}$ is the maximum allowed size of $\delta(x,y)$.

B. Foreground Detection

In this module estimated background and foreground mask images are used as an input for further processing. Thus, we use grayscale image sequences as input. Elements of the scene and the sizes of the traffic objects (vehicles and pedestrians) are unknown. The Foreground detection is done by using accumulative difference method, which is change-detection based on subtraction of a background image. It is necessary to update the background image frequently in order to guarantee reliable object detection. The basic idea in background adaptation is to integrate the new incoming information into the current background image using a Kalman filter:

$$B_{(t+1)} = B_t + [a_1 * 1 - M_t + a_2 * M_t] D_t, \quad (3)$$

where B_t represents the background model at time t, D_t is the difference between the present frame and the background model, and M_t is the binary moving objects hypothesis mask. The gain a_1 *and* a_2 are based on an estimate of the rate of change of the background. The larger it is, the faster new changes in the scene are updated to the background frame. In our approach, $a_1 = 0.1$ *and* $a_2 = 0.01$, they are kept small and the update process based on Eq.(3) is only intended for adapting to slow changes in overall lighting.

$$M_t(x) = \begin{cases} 1, & if\ |D_t(x)| > T_t \\ 0, & Otherwise \end{cases}. \quad (4)$$

Foreground detection is started by computing a pixel based absolute difference between each incoming frame $I(x,y)$ and an adaptive background frame $B_t(x,y)$. The pixels are assumed to contain motion if the absolute differences exceed a predefined threshold level.

$$F(x,y) = \frac{|I(x,y) - B_t(x,y) - \mu|}{\sigma} > T \quad (5)$$

As a result, a binary image is formed where active pixels are labeled with a "1" and non-active ones with a "0". With the updated background image strategy using Kalman filter, we get the better foreground detection result. This is a simple, but efficient method to monitor the changes in active during a few consecutive frames. Those pixels which tend to change their activity frequently are masked out from the binary image representing the foreground detection result.

C. Shadow Removal

Shadows appear as surface features, when they are caused by the interaction between light and objects. This may lead to problems in scene understanding, object segmentation, tracking, recognition, etc. Because of the undesirable effects of shadows on image analysis, much attention was paid to the area of shadow detection and removal over the past decades and covered many specific applications such as traffic surveillance. In this paper, 8-neighborhood gray clustering method is used to define the precise shadow and remove it. The mean clustering threshold and the initial cluster seed of the gray are calculated by the following equations.

$$T_i = (1/3)\max(G(x,y) - u_i)^2, \quad (6)$$

where $G(x,y)$ is a gray value of the pixel in $I(x,y)$, u_i is the mean of $G(x,y)$. The initial seed C_i locates in the centre of $G(x,y)$.The clustering starts from the seed C_i, and the point P_i is examined in turn. If at least one point in the 8-neighborhood of P_i has been marked as a shadow region, standard deviation of P_i is calculated by the following equation.

$$\delta P_i = (P_i(x,y) - u_i)^2, \quad (7)$$

where $P_i(I(x,y))$ is the gray value of I(x,y). If $\delta P_i < T$, P_i must be shadow point; otherwise, the point need not be marked. The point P_i is checked constantly until no new point is marked. At last all the marked shadow points are removed.

D. Noise removal

In regular practice due to the camera noise and irregular object motion, there are some noise regions existed in both the object and background regions. In our method we have incorporated Gaussian noise with the acquired image and propose a solution to see how the background subtraction module would behave while the traditional background algorithms are not providing the significant results. The focus is on the background subtraction module because image noise mostly impacts the foreground extraction process. If the foreground objects are not detected well, the rest of the modules will possibly fail at their tasks.

In the proposed method, after finding the foreground object, noise is estimated and modeled using the following algorithm.

Step 1: Convert RGB image of $F(x,y)$ into gray scale image.

Step 2: Motion blurring can be estimated in a spatially linear invariant system under certain conditions. If we assume the object translates at a constant velocity *V* during the exposure time T with angle α, then the distortion $d = VT$ and define the point spread function (PSF) as

$$h(x,y) = \begin{cases} \frac{1}{d}, & 0 \le |x| \le d\cos\alpha, y = d\sin\alpha \\ 0 & otherwise \end{cases}, \quad (8)$$

The motion blur can be described mathematically as the result of a linear filter

$$g(x,y) = h(x,y) * F(x,y) + \eta(x,y), \quad (9)$$

where $g(x,y)$ denotes the blurred image, $f(x,y)$ denotes the original image, $\eta(x,y)$ denotes additive noise and * represents 2-D convolution.

Step 3: Dividing the Fourier transform of the PSF into the transform of the blurred image, and performing an inverse FFT, reconstruct the image without noise.

$$\Im(f(x,y)) * \Im(PSF) = \Im(g(x,y)), \quad (10)$$

$$f(x,y) = \Im^{-1}\left[\frac{\Im(g(x,y))}{\Im(PSF)}\right], \quad (11)$$

where $f(x,y)$ is the original image and $g(x,y)$ acquired blurred image. $\Im(.)$ and $\Im^{-1}(.)$ represent Fourier and inverse Fourier transform.

In order to estimate the point-spread-function (PSF) we make use of autocorrelation functions $K(n)$ of an M pixel image line l, which is defined as,

$$K(n) = \sum_{i=-M}^{M} l(i+n)l(i) \qquad n \in [-M, M], \quad (12)$$

where $l(i) = 0$ outside the image line range. The above Equation describes how pairs of pixels at particular displacements from each other are correlated. It is high where they are well correlated and low where poorly correlated. For a normal image, the $K(n)$ will be some function of distance from the origin plus random noises. But for a motion blurred image, the $K(n)$ will decline much more slowly in the direction of the blur than in other directions.

Step 4: The Noise to Signal Power Ratio was computed using the following equations:

PSNR is measured in decibels. It's defined as

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2, \quad (13)$$

Where i, j are the width and height of the frame, respectively, in pixels. The PSNR is defined as

$$PSNR = 20\log_{10}\left(\frac{MAX_I^2}{\sqrt{MSE}}\right), \quad (14)$$

Step 5: Apply Wiener filter with θ and d to deblur the image.

Step 6: The Wiener filtering is employed on the resultant Autocorrelation matrices. Wiener filtering minimizes the expected squared error between the restored and perfect images. A simplified Wiener filter is as follows:

$$\hat{F}(x,y) = \left[\frac{1}{H(x,y)}\frac{|H(x,y)|^2}{|H(x,y)|^2 + S_\eta(x,y)/S_f(x,y)}\right]G(x,y), \quad (15)$$

where $H(x,y)$ is the degradation function

$$|H(x,y)|^2 = H^*(x,y)H(x,y), \quad (16)$$

$H^*(x,y)$ is the complex conjugate of $H(x,y)$, $S_\eta(x,y) = |N(x,y)|^2$ represents PSF of the noise, $S_f(x,y) = |F(x,y)|^2$ denotes the PSF of the original image $S_\eta(x,y)/S_f(x,y)$ is called the noise to signal power ratio.

If noise power spectrum is zero, the Wiener filter reduces to a simple inverse filter. If K is large compared with $|H(x,y)|$, then the large value of the inverse filtering term $1/H(x,y)$ is balanced out with the small value of the second term inside the brackets.

Step 7: The Lucy-Richardson filtering on the resultant Autocorrelation matrices. Using Lucy-Richardson method the image is modeled by maximizing the likelihood function gives an equation that is satisfied when the following iteration converges

$$\hat{f}_{k+1}(x,y) = \hat{f}_k(x,y)\left[h(-x,-y) * \frac{g(x,y)}{h(x,y) * \hat{f}_k(x,y)}\right], \quad (17)$$

where * indicates the convolution. $\hat{f}$ is the estimate of the original image f at step k of the iteration. The iteration starts with $f_0 = g(x,y) * h(x,y)$ and $g(x,y)$ denotes the blurred image, $h(x,y)$ denotes the PSF.

E. Detection

After post-processing, the image is compared with the one of the original frames (usually, the first frame). If the pixels are less than certain threshold, then they are ignored. Otherwise, they are replaced by the pixels of original image. This resulting image will be consisting of the moving object ignoring the background and hence satisfying our requirement.

IV. IMPLEMENTATION AND PERFORMANCE ANALYSIS

This system was implemented on an Intel Pentium IV 280 GHz PC. We have tested the system on image sequences on different scenarios like traffic junction intersection, highways etc. Real life traffic video sequence is used to demonstrate the vehicle tracking from traffic video sequences using the proposed framework. All the videos chosen for vehicle tracking have same light intensity and have been taken during day time. We convert the colour video frames to gray scale images.

Automatic monitoring visual surveillance system implementation needs to detect vehicles using automatic background extraction. Background subtraction is the main step for vehicle detection. Fig. 4 shows number of successive frames that are used to extract the background. Digital camera used to take shots. The camera placed over the highway directly. It shots eight frames per second.

Figure 4. Number of successive frames that are used for preprocessing.

Estimated background using Median Filter for n = 12. This is shown in Fig.5

Figure 5. Estimated background image.

After applying the threshold, φ to the absolute difference we got the binary moving objects hypothesis mask which is shown in Fig. 6. With no noise information, the Wiener and other filters do a poor job at realizing the original, non-degraded, image. However, the Lucy-Richardson filter works really well, despite having no information about the noise in the image. With noise information, the Wiener filter gives better result than other filters. Fig 7. shows the comparison of PSNR value with varies filters.

Figure 6. Binary moving objects hypothesis mask.

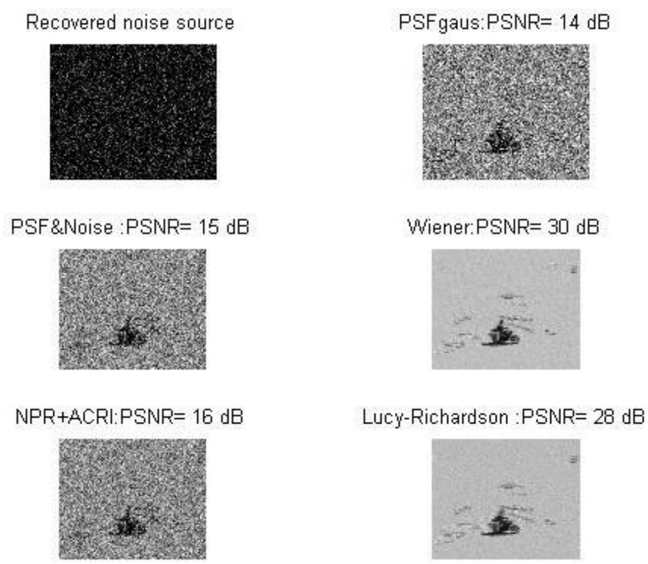


Figure 7. Comparison of PSNR value with a mixture of filters.

Table I shows the result of a mixture of filters used in the Noise removal module. Fig 8 shows the Comparative Analysis of mixture of filters using chart.

Table I. Result of Nose removal module in decibel.

Object used for analysis	Gaussian	NPR	Wiener	NPR + Autocorrelatio	L-R
car	13	13	25	13	24
bus	12	13	24	13	22
motorcycle	14	15	30	16	28
Lorry	10	25	24	23	25
Truck	14	20	26	20	24

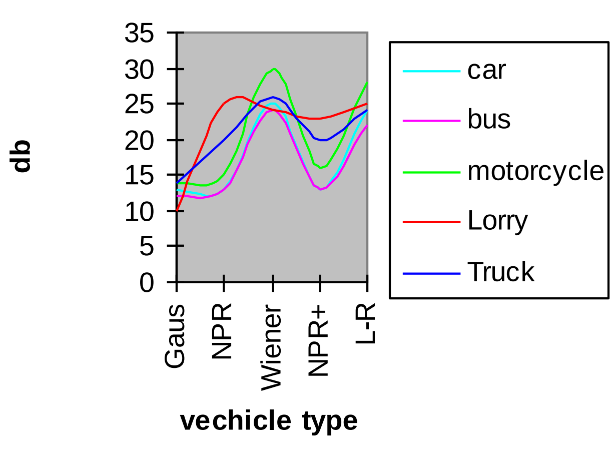


Figure 8. Comparative Analysis of a mixture of filters.

Fig.9 shows the foreground detected objects obtained after background subtraction, shadow and noise. The automatic background extraction results are very good and promising. The most effective parameters that are playing a main role for automatic background extraction are the threshold level. This threshold is used to extract the moving vehicles from the background. Matlab built-in function has been employed for the evaluation of the threshold.

Table II gives vehicle detection results. The background is subtracted from the current image then the resulted image is filtered to get moving vehicles only. By using this technique most of vehicles are detected. Moving vehicles are detected easily after background is subtracted. Performance analysis is shown in Fig.10.

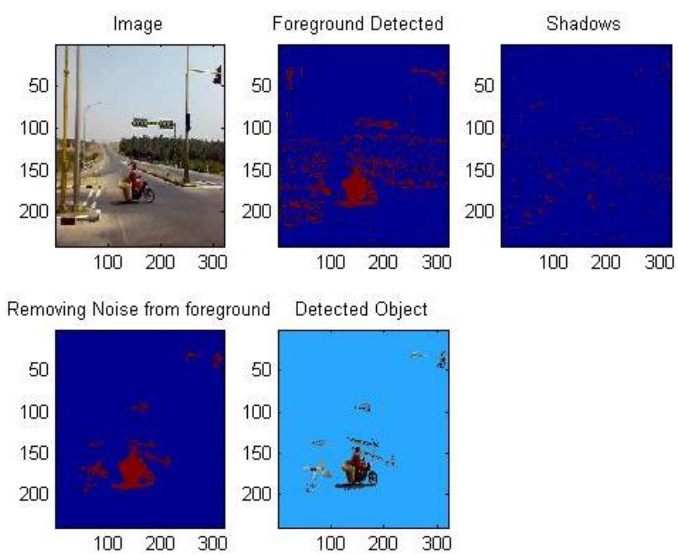


Figure 9. Sequence of steps in the foreground object detection process.

TABLE II. The results for vehicle detection after noise removal

Type of Vehicle	Actual Number of vehicles	Detected Vehicles	Rate %
Car	30	28	93
Motor cycle	30	29	97
Bus	30	27	90
Lorry	30	27	90
Truck	30	28	93

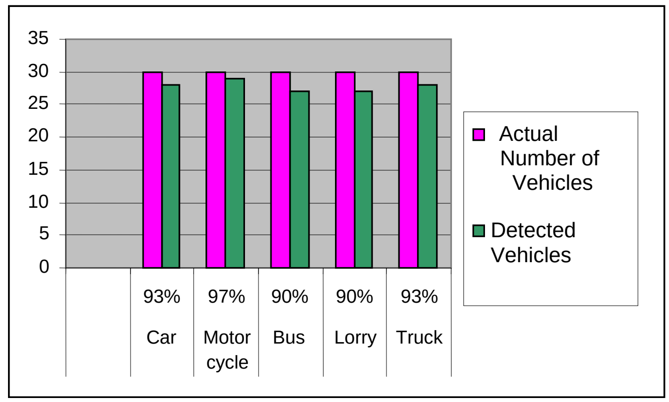


Figure.10 Performance analysis of the proposed method in detection different vehicles in the noisy traffic environment.

V. CONCLUSION

The experimental results of using this approach lead to detect moving vehicles efficiently. This algorithm has been implemented and evaluated experimentally using natural traffic images. It gives promising and effective results where an average vehicle detection rate was around 94%. In this approach the background subtraction and edge detection are used. It is mainly focused on the autocorrelation method. And then an adaptive algorithm is applied to autocorrelation. PSNR is used for the evaluation to measure the performance of the filters. It could be improved and used as a basis for automatic traffic monitoring. Failure detection resulted from occluding large vehicles with small ones and the far moving vehicles that appear as a point in the image. These difficulties could be solved in the future work.

ACKNOWLEDGMENT

Authors thank their family members and children for their continuous support and consent encouragement to do this research work successfully.

REFERENCES

[1] M. Cristani, M. Bicegi, and V. Murino, “Integrated Region- and Pixel-based Approach to Background Modeling”, Proceedings of the MOTION, 2002.

[2] H. Eng, J. Wang, A. Kam, and W. Yau, “Novel Regionbased Modeling for Human Detection within High Dynamic Aquatic Environment,” Proceedings on CVPR, 2004.

[3] P. KaewTraKulPong and R. Bowden, “An Improved Adaptive Background Mixture Model for Real-time Tracking with Shadow Detection,” In Proc. 2nd European Workshop on Advanced Video Based Surveillance Systems, 2001.

[4] M. Harville, G. Gordon, and J. Woodfill, "Foreground Segmentation Using Adaptive Mixture Models in Color and Depth," Proc. ICCV Workshop Detection and Recognition of Events in Video, July 2001.

[5] C. Stauffer and W.E.L. Grimson, "Adaptive Background Mixture Models for Real-Time Tracking," Proc. Conf. Computer Vision and Pattern Recognition, vol. 2, pp. 246-252, June 1999.

[6] S.J. McKenna, Y. Raja, and S. Gong, "Object Tracking Using Adaptive Color Mixture Models," Proc. Asian Conf. Computer Vision, vol. 1, pp. 615-622, Jan. 1998.

[7] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, "Detecting Moving Objects, Ghosts, and Shadows in Video Streams," IEEE Trans. on PAMI, 25: (10), October 2003.

[8] Connell, J., "Detection and Tracking in the IBM PeopleVision System," IEEE ICME, June 2004.

[9] Dongxiang Zhou, Hong Zhang and Nilanjan Ray, "Texture Based Background Subtraction," *Proc. IEEE Int. Conf. on Information and Automation*, Zhangjiajie, China, pp. 601-605, 2008.

[10] A. McIvor, "Background subtraction techniques," in *Proc. Image Video Computing*, pp. 147-153, 2000.

[11] Y. Ivanov, A. Bobick, and J. Liu, "Fast lighting independent background subtraction," *Int. Journal on Comp. Vision*, vol. 37, no. 2, pp. 199–207, Jun. 2000.

[12] Liyuan Li, Weimin Huang, Irene Yu-Hua Gu and Qi Tian, "Statistical Modeling of Complex Backgrounds for Foreground Object Detection," *IEEE Transactions on Image Proc.*, vol. 13, no. 11, pp. 1459-1472, Nov 2004.

[13] A. Elgammal, D. Harwood, and L.S. Davis, "Non-Parametric Model for Background Subtraction," Proc. IEEE Int'l Conf. Computer Vision '99 FRAME-RATE Workshop, 1999.

[14] M. Sankari, Dr. C. Meena "Adaptive Background Estimation and object detection applying in Automated Visual surveillance" International Journal of Computer science and Information Security July 2010, Vol. 8 No. 4.

[15] Peng Lei, "Adaptive median filtering", Machine vision, 140.429 Digital image processing.

AUTHORS' PROFILE

Mrs. M. Sankari received her B.Sc. and M.Sc. degrees in Computer science from Bharathidasan University in 1988 and 1990, respectively. She has completed her Master of Philosophy degree in Computer science from Regional Engineering College, Trichy in 2000. Presently, she is a Assistant Professor & Head in the department of MCA at NIET. She is pursuing her Ph.D. degree in Computer science at Avinashilingam University for women, Coimbatore, India. She has published several technical papers in IEEE/IET conferences and Journal. Her field of research includes Computer vision, Pattern recognition, Analysis of algorithms, Data structure, Computer graphics and multimedia.

Dr. C. Meena received her B.Sc (Physics), and Master of Computer Applications degrees from Madurai Kamaraj University in 1987 and 1990, respectively. She has completed her Ph.D. degree in Computer science from Bharathiyar University in 2006. Presently, she is Head, Computer Centre at Avinashilingam University, Coimbatore, India. She is guiding for funded Research Project in UGC and Naval Research Board. She has presented international/national journals. She has published various technical papers in IEEE/IET conferences and national level conferences. Her field of research includes Image processing, Computer vision and Pattern recognition.

Improvement of Brain Tissue Segmentation Using Information Fusion Approach

Lamiche Chaabane
Department of Computer Science
University of M'sila
Algeria

Moussaoui Abdelouahab
Department of Computer Science
University of Setif
Algeria

***Abstract*— The fusion of information is a domain of research in full effervescence these last years. Because of increasing of the diversity techniques of images acquisitions, the applications of medical images segmentation, in which we are interested, necessitate most of the time to carry out the fusion of various data sources to have information with high quality. In this paper we propose a system of data fusion through the framework of the possibility theory adapted for the segmentation of MR images. The fusion process is divided into three steps : fuzzy tissue maps are first computed on all images using Fuzzy C- Means algorithm. Fusion is then achieved for all tissues with a fusion operator. Applications on a brain model show very promising results on simulated data and a great concordance between the true segmentation and the proposed system.**

Keywords- Information fusion; possibility theory; segmentation; FCM; MR images.

I. INTRODUCTION

Recent technical advances have led to the multiplication of imaging systems, which are often used for observing a phenomenon from different points of view. They provide a large amount of information that must a whole in order to draw correct conclusions. This development has made data fusion in image processing an important step, now well recognized, in modern multi source image analysis. In medical imaging in particular the clinician may use images issued from different sources, each of them highlighting specific properties of tissues and pathologies. They may be images acquired with a single imaging technique using different acquisition parameters (for instance multi echo MRI), or images were obtained from several imaging techniques (for instance anatomical MR imaging combined with functional PET imaging). The association of such images allows the medical expert to confirm and to complete his diagnosis because it can arrive as none the images available contains a sufficient information separately. In most data fusion problems, the images to be combined are partly redundant as they represent the same scene, and partly complementary as they may highlight different characteristics [1].

Typically, none of the images provide a completely decisive and reliable information. In addition, the information is often imprecise and uncertain, and these characteristics are inherent to the images; due to observed phenomenon, sensors, numerical reconstruction algorithms and resolution etc. The aim of data fusion techniques is therefore to improve the decision by increasing the amount of global information, while decreasing its imprecision and uncertainty by making us of redundancy and complementarities [1].

Some mathematical models were discussed in the literature for the modelization of both uncertainty and imprecision. Traditionally probabilities theory was the primary model used to deal with uncertainty problems, but they suffer from drawbacks which are still a matter of discussion. Whereas the Dempster-Shafer theory also allows to representing these two natures of information using functions of mass but the set of operators used by this theory in fusion step is very restricted. Alternative to this approach is the possibility theory where uncertainty and imprecision are easily modeled, in this article we will focus on this last one for two essential reasons : this theory allows to combining information coming from various sources by the use a wide range of available combination operators. In addition, this theory seems to us the most adapted to the considered problem in the modeling step [1][2].

We present in this article, a fuzzy information fusion framework for the automatic segmentation of human brain tissues using T2- weighted (T2) and proton density (PD) images. This framework consists of the computation of fuzzy tissue maps in both images by means of Fuzzy C-Means algorithm, the creation of fuzzy maps by a combination operator and a segmented image is computed in decision step.

The organization of the paper as follows : In section II, some previous related works are presented. In section III, we briefly outline the principals of possibility theory reasoning. Section IV discusses the architecture of the fusion system. Steps of fusion in possibility theory are explained in Section V. In section VI the traditional FCM algorithm is briefly reviewed. Section VII presents our proposed approach. Some experiment results using two routing MR sequences T2 and PD feature images are shown in section VIII and section IX contains conclusions and addresses future works.

II. RELATED WORKS

Many works have been done in the field of fuzzy information fusion in the literature. A brief review of some of them is presented in this section. Waltz [6] presented three basic levels of image data fusion : pixel level, feature level and decision level, which correspond to three processing architectures. I. Bloch [1] have outlined some features of Dempster-Shafer evidence theory, which can very useful for medical image fusion for classification, segmentation or recognition purposes. Examples were provided to show its ability to take into account a large variety of situations. Registration-based methods are considered as pixel-level fusion, such as MRI-PET (position emission tomography) data fusion [7]. Some techniques of knowledge-based segmentation can be considered as the feature-level fusion such as the methods proposed in [11].

Some belief functions, uncertainty theory, Dempster-Shafer theory are often used for decision-level fusion such as in [9]. In [12], I. Bloch proposed an unified framework of information fusion in the medical field based on the fuzzy sets, allow to represent and to process the numerical data as well as symbolic systems, the fuzzy sets theory is applied to three levels: at the low level to treat the basic numerical information contained in the images, as well as possible ambiguity between the classes; on the level object, to represent objects or structures in the images such as a fuzzy objects. at the higher level, to take into account a structural information and some characteristics as the distance, adjacency, and the relative position between objects.

V. Barra and J. Y. Boire [4] have described a general framework of the fusion of anatomical and functional medical images. The aim of their work is to fuse anatomical and functional information coming from medical imaging, the fusion process is performed in possibilistic logic frame, which allows for the management of uncertainty and imprecision inherent to the images. They particularly focus on the aggregation step with the introduction of a new class of operators based on information theory and the whole process is finally illustrated in two clinical cases : the study of Alzheimer's disease by MR/SPECT fusion and the study of epilepsy with MR/PET/SPECT. The obtained results was very encouraging.

V. Barra and J. Y. Boire [10] proposed a new scheme of information fusion to segment intern cerebral structures. The information is provided by MR images and expert knowledge, and consists of constitution, morphological and topological characteristics of tissues. The fusion of multimodality images is used in [8]. In [3], the authors have presented a framework of fuzzy information fusion to automatically segment tumor areas of human brain from multispectral magnetic resonance imaging (MRI) such as T1-weighted, T2-weighted and proton density (PD) images; in this approach three fuzzy models are introduced to represent tumor features for different MR image sequences. They allow to create corresponding fuzzy feature space of tumor. All the t-norm or fuzzy intersection operators can be used as fusion operators for this fuzzy features. the geometric mean is chosen using experiments allowing us to take correctly into account the three fuzzy spaces in a simple way. The fuzzy region growing is used to improve the fused result.

Maria del C and al [5] proposed a new multispectral MRI data fusion technique for white matter lesion segmentation, in that a method is described and comparison with thresholding in FLAIR images is illustrated.

III. THE POSSIBILITY THEORY

Possibilistic logic was introduced by Zadeh (1978) following its former works in fuzzy logic (Zadeh, 1965) in order to simultaneously represent imprecise and uncertain knowledge. In fuzzy set theory, a fuzzy measure is a representation of the uncertainty, giving for each subset Y of the universe of discourse *X* a coefficient in [0,1] assessing the degree of certitude for the realization of the event *Y*. In possibilistic logic, this fuzzy measure is modeled as a measure of possibility Π satisfying:

$$\Pi(X) = 1 \quad et \quad \Pi(\phi) = 0$$

$$(\forall(Y_i))\Pi(\bigcup_i Y_i) = \sup_i \Pi(Y_i)$$

An event Y is completely possible if $\Pi(Y) = 1$ and is impossible if $\Pi(Y) = 0$. Zadeh showed that Π could completely be defined from the assessment of the certitude on each singleton of *X*. Such a definition relies on the definition of a distribution of possibility π satisfying :

$$\pi : X \to [0,1]$$
$$x \to \pi(x) / \sup_x \{\pi(x) = 1\}$$

Fuzzy sets F can then be represented by distributions of possibility, from the definition of their characteristic function μ_F :

$$(\forall x \in X) \mu_F(x) = \pi(x)$$

Distributions of possibility can mathematically be related to probabilities, and they moreover offer the capability to declare the ignorance about an event. Considering such an event A (e.g., voxel v belongs to tissue T, (where v is at the interface between two tissues), the probabilities would assign $P(A) = P(\overline{A}) = 0.5$, whereas the possibility theory allows fully possible $\Pi(A) = \Pi(\overline{A}) = 1$. We chose to model all the information using distributions of possibility, and equivalently we represented this information using fuzzy sets [18].

The literature classically distinguishes three modes for combination of uncertainty and imprecise information in a possibility theory framework [23] :

The conjunction: gather the operators of t-norms (fuzzy intersection), this mode of combination must be used if measurements are coherent, i.e. without conflict.

The compromise: gather the median operator and some average operators, it must be used when measurements are in partial conflict.

The Disjunction: gather the operators of t-conorms (fuzzy union), it must be used when measurements are in disaccord, i.e. in severe conflict.

In introduction, we underlined the inopportunity to combining information in a fixed mode: if observations are in accord, it is legitimate to combine them in a conjunctive mode or compromise in order to extract a more relevant information. But if a serious conflict appears, it is better to combining in a disjunctive mode. For example, if two measurements of the same parameter prove completely different, it is not judicious to make an average of it, better is worth to say than one or the other is true [24].

IV. The Fusion process and Type of Architectures

A general information fusion problem can be stated in the following terms : given l sources S_1, $S_2,\ldots S_l$ representing heterogeneous data on the observed phenomenon, take a decision d_i on an element x, where x is higher level object extracted from information, and Di belongs to a decision space $D=\{d_1, d_2, d_3,\ldots, d_n\}$ (or set of hypotheses). In numerical fusion methods, the information relating x to each possible decision d_i according to each source S_j is represented as a number M_{ij} having different properties and different meanings depending on the mathematical fusion framework. In the centralized scheme , the measures related to each possible decision i and provided by all sources are combined in a global evaluation of this decision, taking the form, for each i : $Mi = F(M_{i1}, M_{i2}, M_{i3}, \ldots, M_{in})$, where F is a fusion operator. Then a decision is taken from the set of M_i, $1\leq i\leq n$. in this scheme, no intermediate decision is taken and the final decision is issued at the end of the processing chain. In decentralized scheme decisions at intermediate steps are taken with partial information only, which usually require a difficult control or arbitration step to diminish contradictions and conflicts [2][4].

The three-steps fusion can be therefore described as :

- Modeling of information in a common theoretical frame to manage vague, ambiguous knowledge and information imperfection. In addition, in this step the M_{ij} values are estimated according to the chosen mathematical framework.
- Combination : the information is then aggregated with a fusion operator F. This operator must affirm redundancy and manage the complementarities and conflicts.
- Decision : it is the ultimate step of the fusion, which makes it possible to pass from information provided by the sources to the choice of a decision d_i.

V. Fusion in Possibility Theory

A. Modeling Step

In the framework of possibility theory and fuzzy sets [13][14[15], the M_{ij}'s represent membership degrees to a fuzzy set or possibility distribution π , taking the form for each decision d_i and source S_i :. $M_{ij} = \pi_j(d_i)$.

B. Fusion step

For the aggregation step in the fusion process, the advantages of possibility theory rely in the variety of combination operators, which must affirm redundancy and manage the complementarities. And may deal with heterogeneous information [16][17][18]. It is particular interest to note that, unlike other data fusion theories like Bayesian or Dempster-Shafer combination, possibility theory provides a great flexibility in the choice of the operator, that can be adapted to any situation at hand [2].

C. Decision step

Is usually taken from maximum of memberships values after the aggregation step. Many constraints can be added to this decision, typically for checking for the reliability of the decision (is he obtained value high enough?) or for the discrimination power of he fusion (is the difference between the two highest values high enough ?) [2].

VI. The FCM Algorithm Clustering

Clustering is a process of finding groups in unlabeled dataset based on a similarity measure between the data patterns (elements) [12]. A cluster contains similar patterns placed together. The fuzzy clustering technique generates fuzzy partitions of the data instead of hard partitions. Therefore, data patterns may belong to several clusters, having different membership values with different clusters. The membership value of data pattern to a cluster denotes similarity between the given data pattern to the cluster. Given a set of N data patterns $X=\{x_1, x_2, x_3, \ldots, x_n\}$ the Fuzzy C-Means (FCM) clustering algorithm minimizes the objective function [26][27]:

$$J(B,U,X) = \sum_{i=1}^{C}\sum_{j=1}^{N} u_{ij}^{m} d^{2}(x_j, b_i) \quad (1)$$

Where x_j is the j-th P-dimensional data vector, b_i is the center of cluster i, u_{ij} is the degree of membership of x_j in the j-th cluster, m is the weighting exponent $d^2(x_j,b_i)$ is the Euclidean distance between data x_j and cluster center b_i.

The minimization of objective function $J(B,U,X)$ can be brought by an iterative process in which updating of membership u_{ij} and the cluster centers are done for each iteration.

$$u_{ij} = \left[\sum_{k=1}^{C}\left(\frac{d^2(x_j,b_i)}{d^2(x_j,b_k)}\right)^{2/(m-1)}\right]^{-1}. \quad (2)$$

$$b_i = \frac{\sum_{k=1}^{N} u_{ij}^{m} x_k}{\sum_{k=1}^{N} u_{ik}^{m}}. \quad (3)$$

Where :

$$\forall i \in \{1..C\},\quad \forall j \in \{1..N\} \qquad \begin{cases} u_{ij} \in [0,1] \\ 0 < \sum_{i=1}^{N} u_{ij} < N \end{cases}. \quad (4)$$

$$\forall j \in \{1..N\} \quad \sum_{i=1}^{C} u_{ij} = 1. \quad (5)$$

The algorithm of the FCM consists then of the reiterated application of (2) and (3) until stability of the solutions.

VII. PROPOSED METHOD

In this section we propose a framework of data fusion based on the possibility theory which allows the segmentation of MR images. The operation is limited to the fusion of T2 and PD images. Then information to combine are thus homogeneous and the scheme of our proposed fusion system as shown in figure 1 below:

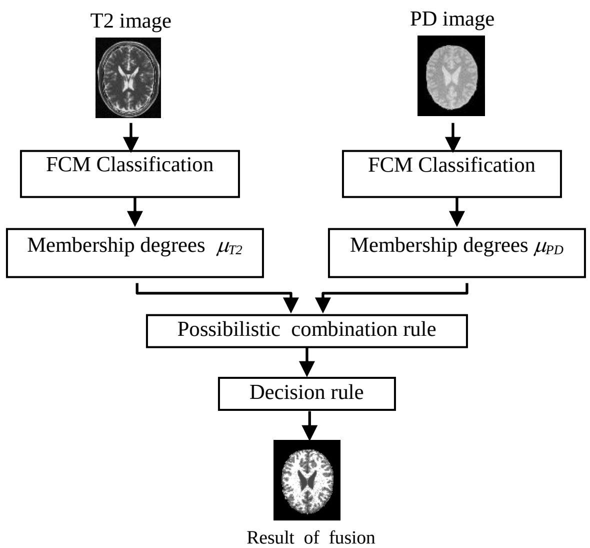


Figure 1. Scheme of the proposed fusion system.

If it is supposed that these images are registered, our approach of fusion consists of three steps:

A. *Modeling of the data*

In this phase the choice of the fuzzy framework is retained to modeling information resulting from the various images. More precisely, MR images are segmented in C = 4 classes using the FCM algorithm described in section VI. For each MR image I, C distributions of possibility π_T^I, $1 \leq T \leq C$ are then obtained, and are represented by memberships of the pixels to the classes.

B. *Combination*

The aggregation step is most fundamental for a relevant exploitation of information resulting from the images I_{T2} , I_{PD}. The operator must combine for a given tissue T, the distributions of possibility π^{T2} and π^{PD}, by underlining the redundancies and managing ambiguities and complementarities between the T2-weighted and proton density images.

1) Choice of an operator : One of the strengths of the possibility theory is to propose a wide range of operators for the combination of memberships. I. Bloch [20] classified these operators not only according to their severe (conjunctive) or cautious (disjunctive) nature but also with respect to their context-based behavior. Three classes were thus defined:

- Context independent and constant behavior operators (CICB);

- Context independent and variable behavior operators (CIVB);

- Context dependent operators (CD).

For our T2/PD fusion, we chose a (CICB) class of combination operators because in the medical context, both images were supposed to be almost everywhere concordant, except near boundaries between tissues and in pathologic areas [20]. Three operators (minimum, maximum, and arithmetic mean) of this class who does not need any parameter were tested related to the fusion of MR images acquired in weighting T2, PD. They were carried out on a range of 70 slices of Brain1320 volume of Brainweb[1] If π_T^{T2} , π_T^{PD} are the possibility distributions of tissue T derived from T2 and PD maps , then the fused possibility as defined for any gray level v as :

The minimum operator: $\pi_T(v) = Min(\pi_T^{T2}(v), \pi_T^{PD}(v))$

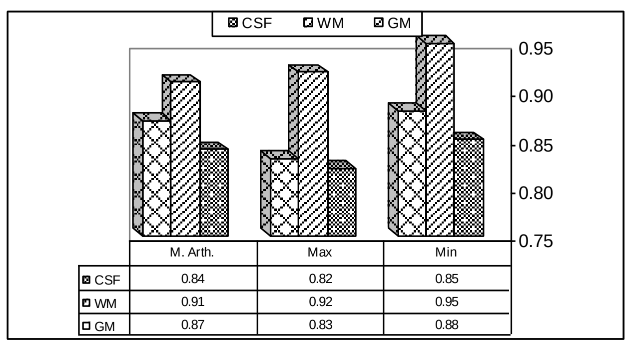
The maximum operator: $\pi_T(v) = Max(\pi_T^{T2}(v), \pi_T^{PD}(v))$

The arithmetic mean operator: $\pi_T(v) = (\pi_T^{T2}(v) + \pi_T^{PD}(v))/2$

These operators are compared with the reference result using the coefficient DSC[2]. Which measures the overlap between two segmentations S1 and S2. It is defined as:

$$DSC(S1, S2) = 2.card(S1 \cap S2)/card(S1 + S2)$$

The results of these tests are shown on figure 2:



	M. Arth.	Max	Min
CSF	0.84	0.82	0.85
WM	0.91	0.92	0.95
GM	0.87	0.83	0.88

Figure 2. Comparison of the operators by the DSC measurement.

The results drawn up in the figure 2 show the predominance of the minimum operator compared to the maximum operator and the arithmetic mean operator. Thus we will retaining this operator for our study.

C. *Decision*

A segmented image was finally computed using all maps of different tissues T, $1 \leq T \leq C$. So certain theories make it possible to consider several types of decision, the theory of the possibilities proposes only the rule of the maximum of

[1] http://www.bic.mni.mcgill.ca/brainweb/
[2] Dice Similarity Coefficient

possibility. We thus retain this one and assign each pixel to the tissue for which it has the greatest membership.

The general algorithm of our system is .

General algorithm

Modeling of the image

For i **in** {T2,PD} **do**

FCM (i) { *Computation of membership degrees for both images*}

End For

Fusion

Possibilistic fusion {*Between each class of T2 image and the same one of PD image* }

Decision

Segmented image

It should be noted that the stability of our system depend to the stability of the algorithm used in the modeling step[26].

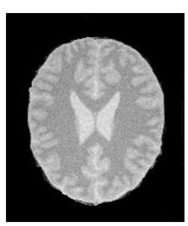
VIII. Results and Disccusion

Since the ground truth of segmentation for real MR images is not usually available, it is impossible to evaluate the segmentation performance quantitatively, but only visually. However, Brainweb provides a simulated brain database (SBD) including a set of realistic MRI data volumes produced by an MRI simulator. These data enable us to evaluate the performance of various image analysis methods in a setting where the truth is known [30][31][32].

to have tests under realistic conditions, three volumes were generated with a thickness of 1 mm and a level of noise of 0%, 3% and 5%. We fixed at 20% the parameter of heterogeneity.

The results of each step of fusion are presented on a noisy 90th brain only slice is shown in figure 3. This noisy slice was segmented into four clusters: background, CSF, white matter, and gray matter using FCM algorithm, however the background was neglected from the viewing results.

Simulated MR T2 image



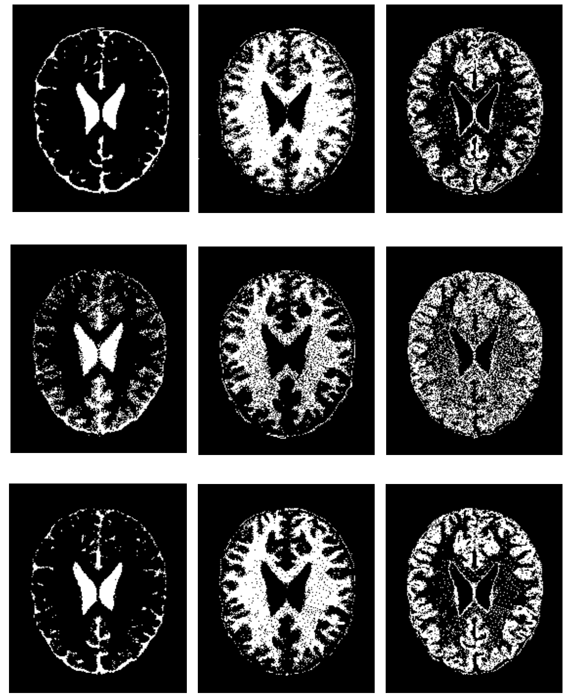
Simulated MR PD image

(a)

(b)

(c)

(d)



CSF WM GM

Figure 3. (a) Simulated T2, PD images illustrate the fusion. (b) Maps of CSF, WM and GM obtained by FCM for T2 image. (c) Maps of CSF, WM and GM obtained by FCM for PD image . (d) Maps of CSF, WM and GM obtained by proposed system.

The results of final segmentation are shown in figure 4 below.

(a) (b) (c)

Figure 4. Segmentation results. (a) T2 segmented by FCM (b) PD segmented by FCM, (c) Image of fusion

The CSF map of PD image is improved significantly by fusion within the noise levels 0% – 5%.

The WM fused map is strongly improved compared to that obtained by the PD only, but This improvement is small compared to that obtained by the segmentation of T2 only.

Information in GM fused map reinforced in area of agreement (mainly in the cortex). and the fusion showed a significant improvement and reduces the effect of noise in images.

These remarks demonstrate the superior capabilities of the proposed approach compared to the taking into account of only one weighting in MR image segmentation.

The performances of our system led us to reflect on the validity of the segmentation obtained. It appeared to us to measure and quantify the performances of our segmentation of the whole of brain. Measurement used is the DSC coefficient described in section 7 and the results are reported in figures 5, 6 and 7.

The graphics of figures 5, 6 and 7 underline the advantages of the fusion of multimodality images within the fuzzy possibilistic framework to improve the results clearly. DSC coefficients obtained by the proposed approach augments the improvement of the segmentation from 2% to 3% for the white matter and from 1% to 3% for the gray matter in T2 image. And from 2% to 12% for the white matter and from 3% to 19% for the grey matter in PD. Image. Moreover one indeed notes that the CSF is improved only compared to the weighting PD, in that the improvement increases by 7% to 21%.

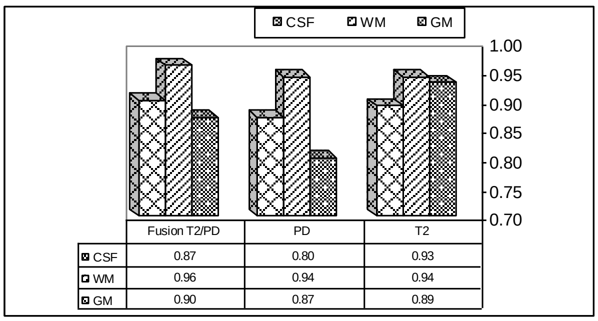


Figure 5. Comparison results between different segmentations with 0% noise

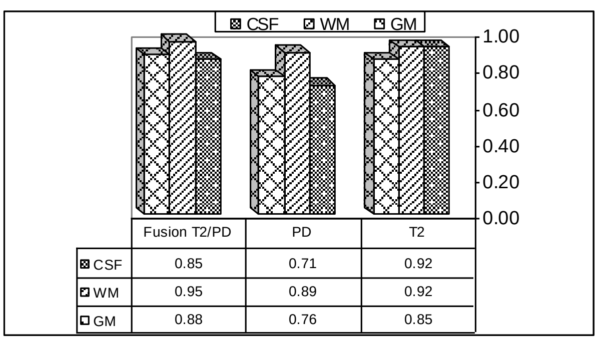


Figure 6. Comparison results between different segmentations with 3% noise

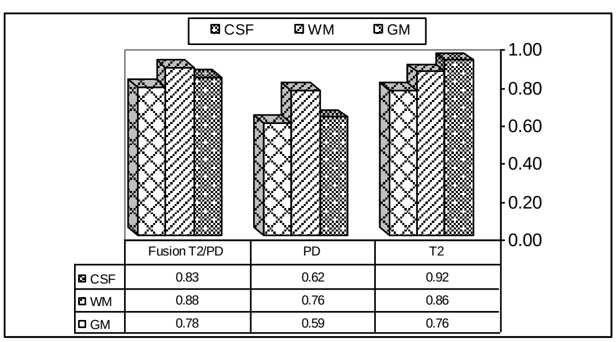


Figure 7. Comparison results between different segmentations with 5% noise

IX. Conclusion

In this article we presented a system of data fusion to segment MR images in order to improve the quality of the segmentation. Since we outlined in here some features of possibility theory, which can be very useful for medical images fusion. And which constitute advantages over classical theories. They include the high flexibility of the modeling offered by possibility theory, taking into account both imprecision and uncertainty and prior information not necessarily expressed as probabilities. The Effectiveness of our system is affirmed by the choice of the model to representing data and the selected operator in a combination step. Results obtained are rather encouraging and underline the potential of the data fusion in the medical imaging field.

As a perspective of this work and on the level of modeling we would wish to integrate other information or new techniques of MR acquisitions and thus to use a more effective and more robust algorithms to representing a data. on fusion level an adaptive operators of fusion are desired for the combination of the data in order to improve the segmentation of the MR images or to detect anomalies in the pathological images.

References

[1] I. Bloch, "Some aspects of Dempster-Shafer evidence theory for classification of multi-modality medical images taking partial volume effect into account," Pattern Recognition Letters, vol. 17, pp. 905–919, 1996.

[2] I. Bloch, and H. Maitre, "Data fusion in 2D and 3D image processing: An overview," X Brazilian symposium on computer graphics and image processing, Brazil, pp. 127–134, 1997.

[3] W. Dou, S. Ruan, Y. Chen, D. Bloyet, J. M. Constans, "A framwork of fuzzy information fusion for the segmentation of brain tumor tissues on MR images," Image and vision Computing, vol. 25, pp. 164–171, 2007

[4] V. Barra and J. Y. Boire, "A General Framework for the Fusion of Anatomical and Functional Medical Images," NeuroImage, vol. 13**,** 410–424, 2001.

[5] D. C. Maria, H. Valdés, J. F. Karen, M. C. Francesca, M. W. Joanna, "New multispectral MRI data fusion technique for white matter lesion segmentation: method and comparison with thresholding in FLAIR images," Eur Radiol, vol. 20, 1684–1691, 2010.

[6] E.D Waltz, the principals and practice of image and spatial data fusion in : Davis L, Hall, James llinas (Eds), Proceedings of the Eight, national Data Fusion Conference ,(Dalls, TX. March 15-17, 1995) Handbook of Multisensor Data Fusion, CRC pess, West Bam Beach, FL,1995, pp. 257–278. (p4–1–4–18).

[7] F. Behloul, M. Janier, P. Croisille, C. Poirier, A. Boudraa, R. Unterreiner, J. C. Mason, D. Revel, "Automatic assessment of myocardial viability based on PET-MRI data fusion," Eng. Med. Biol. Mag., Proc. 20th Ann. Int. Conf. IEEE, vol. 1, pp. 492–495, 1998.

[8] M. Aguilar, R. Joshua, "New fusion of multi-modality volumetric medical imagery," ISIF, pp. 1206–1212, 2002.

[9] E. Lefevre, P. Vannoorenberghe, O. Colot, "About the use of Dempster-Shafer theory for color image segmentation," First international conference on color in graphics and image processing, Saint-Etienne, France, 2000.

[10] V.Barra, and J-Y Boire, "Automatic segmentation of subcortical brain structures in MR images using information fusion," IEEE Trans. on Med. Imaging., vol. 20, n7, pp. 549–558, 2001.

[11] M. C. Clarck, L.O. Hall, D.B. Goldgof, R Velthuizen, F.R. Murtagh, M.S. Silbiger, "Automatic tumor segmentation using knowledge-based techniques," IEEE Trans. Med. Imaging, vol. 17, pp. 187–201, 1998.

[12] I. Bloch, "Fusion of numerical and structural image information in medical imaging in the framework of fuzzy sets," Fuzzy Systems in Medicine, ed. par P. Szczepaniak *et al.*, pp. 429–447, Springer Verlag, 2000.

[13] L. A. Zadeh, “Fuzzy sets,” Information and Control, vol. 8, pp. 338–353, 1965.
[14] L. A. Zadeh, “fuzzy sets as a basic for a theory of possibility,” Int. Jour. of Fuzzy Sets and Systems, vol. 1, pp. 3–28, 1978.
[15] D. Dubois and H. Prade. Fuzzy sets and systems : Theory and Application, New-York: Academic Press, 1980.
[16] D. Dubois and H. Prade, “A review of Fuzzy Set Aggregation Connectives,” Information Sciences, vol. 36, pp. 85–121, 1985.
[17] R. R. Yager, “Connectives and quantifiers in fuzzy sets,” Int. Jour. of Fuzzy sets and systems, vol. 40, pp. 39–75,1991.
[18] D. Dubois and H. Prade, “Combination of information in the framework of possibility theory,” In M. AL Abidi et al., editor, Data Fusion in Robotics and Machine Intelligence: Academic Press, 1992.
[19] V. Barra, Fusion d'images 3D du cerveau : Etude de modèles et Applications. Thèse de doctorat, Université d’Auvergne, 2000.
[20] I. Bloch, “Information combination operators for data fusion : A Comparative Review with Classification,” IEEE Transactions en systems, Man. and Cybernitics, vol. 1, pp. 52–67, 1996.
[21] W. Dou, Segmentation des images multispectrales basée sur la fusion d'information: application aux images IRM. Thèse de doctorat Université CAEN/Basse-Normandi , 2006.
[22] A.Bendjebbour, W.Pieczynski, “Segmentation d'images multisenseurs par fusion évidentielle dans un contexte markovien,” Traitement du Signal, vol. 14, n15, pp. 453–463, 1997.
[23] H.J. Zimmermann. Fuzzy Sets Theory and its Applications. 2nd ed. Boston: Kluwer Academic Publishers, 1991.
[24] S. Deveughelle, B. Dubuisson, “Adaptabilité et combinaison possibiliste: application à la vision multi caméras,” Traitement du Signal, vol. 11, n6, pp. 559–568, 1994.
[25] C.Barillot, J.C.Gee,L.Le Briquer,G.Le Goualher, “Fusion intra et inter individus en imagerie médicale appliquée à la modélisation anatomique du cerveau humain,” Traitement du Signal, vol. 11, n6, 1994.
[26] J. Bezdek, “A convergence theorem for the fuzzy data clustering algorithms,” IEEE Transactions on Pattern Analysis and Machine Intelligence TPAMI, vol. 2, pp. 1–8,1980.
[27] J. Bezdek , L. Hall, L. Clarke, “Review of MR image segmentation using pattern recognition,” Medical Physics, vol. 20, pp. 1033–1048, 1993.
[28] V.Barra, J. H. Boire, Segmentation floue des tissus cérébraux en IRM 3D: une approche possibiliste versus autres méthodes, Rencontres Francophones sur la logique floue et ses applications, Valenciennes, Editions Cépaduès, pp. 193–198, 1999.
[29] I.Bloch, H. Maitre, “Fusion de données en traitement d'images : modèles d'informations et décisions,” Traitement du Signal, vol. 11, n6, pp. 435–446, 1994.
[30] BrainWeb [Online]. Available: www.bic.mni.mcgill.ca/brainweb/
[31] C. A. Cocosco, V. Kollokian, R. K. -S. Kwan, and A. C. Evans, “BrainWeb: Online interface to a 3D MRI simulated brain database,” *NeuroImage*, pt. 2/4, vol. 5, n4, p. S425, 1997.
[32] D. L. Collins, A. P. Zijdenbos, V. Kollokian, J. G. Sled, N. J. Kabani, C. J. Holmes, and A. C. Evans, “Design and construction of a realistic digital brain phantom,” IEEE Trans. Med. Imaging, vol. 17, n3, pp. 463–468, 1998.
[33] I. Bloch, “Fusion d'informations en traitement du signal et des images,” unpublished.

AUTHORS PROFILE

Chaabane Lamiche received his BSc in Computer Science in 1997 from the Department of Computer Science from Ferhat Abbas University, Algeria. He also received Master's degree in Computer Science in 2006 from University of M'sila. He has 4 years experience in teaching. His areas of interests include Data mining and Warehousing, artificial intelligence, Image processing and operational research. His current research interests include the data mining techniques and medical image analysis.

Abdelouahab Moussaoui received his BSc in Computer Science in 1990 from the Department of Computer Science from the University of Science and Technology of Houari Boumedienne (USTHB), Algeria. He also received and MSc in Space Engineering in 1991 from University of Science and Technology of Oran (USTO). He received also a MSc degree in Machine Learning from Reims University (France) since 1992 and Master's degree in Computer Science in 1995 from University of Sidi Bel-abbes, Algeria and PhD degree in Computer Science from Ferhat Abbas University, Algeria. He is IEEE Member and AJIT, IJMMIA & IJSC Referee. His researches are in the areas of clustering algorithms and multivariate image classification applications. His current research interests include the fuzzy neuronal network and non parametric classification using unsupervised knowledge system applied to biomedical image segmentation. He also works from a long time on pattern recognition’s algorithm, complex data mining and medical image analysis.

Image Retrieval using DST and DST Wavelet Sectorization

Dr. H.B.Kekre
Sr. Professor, Computer Engineering Dept.
MPSTME, SVKM's NMIMS (Deemed-to be University)
Vile Parle West, Mumbai, INDIA

Dhirendra Mishra
Associate Professor and PhD. Research scholar
MPSTME, SVKM's NMIMS (Deemed-to be University)
Vile Parle West, Mumbai, INDIA

***Abstract*—The concept of sectorization of the transformed images for CBIR is an innovative idea. This paper introduces the concept of Wavelet generation for Discrete sine transform (DST).The sectorization of the DST transformed images and the DST wavelet transforms has been done into various sector sizes i.e.4,8,12, and 16. The transformation of the images is tried and tested threefold i.e. row wise transformation, column wise transformation and Full transformation. We have formed two planes i.e. plane 1 and plane 2 for sectorization in full transformation .The performance of the all the approaches has been tested by means of the three plots namely average precision-recall cross over point plot, LIRS (Length of initial relevant string of images) plot, LSRR (Length of string to recover all relevant images in the database) plot. The algorithms are analyzed to check the effect of three parameters on the retrieval First the way of transformation (Row, Column, Full), Second the size of sector generated, Third the type of similarity measures used. With the consideration of all these the overall comparison has been performed.**

Keywords-CBIR, Feature extraction; Precision; Recall; LIRS; LSRR; DST; DST Wavelet.

I. Introduction

The digital world Innovations has evolved itself to a very large extent. The result of which has increased the more and more dependency on the digital data and in turn on computer system. The information of any form i.e. multimedia, documents, images etc. everything has got its own place in this digital world. The computer system has been accepted to be the very powerful mechanism to use these digital data, for its secured storage and efficient accessibility whenever required.

Digital Images play a very good role for describing the detailed information about man, money, machine etc. almost in every field. The various processes of digitizing the images to obtain it in the best quality for the more clear and accurate information leads to the requirement of more storage space and better storage and accessing mechanism in the form of hardware or software. As far as the accessing of these images are concerned one needs to have the good mechanism of not only for accessing of the images but also for any other image processing to be done one needs to have the faster, accurate, efficient retrievals of these images. There are various approaches of proposing the methodologies of retrieving the images from the large databases consisting of millions of images stored.

Content Based Image Retrieval (CBIR) [1-4] is one of the evolving fields of image processing. CBIR needs to have the innovative algorithm to extract the perfect features to define identity of an image. It has been researched upon to use content of the image itself to draw out its unique identity. This unique identity can make one to differentiate the images with each other with better and accurate retrieval of images. There are mainly three contents i.e. shape, color and textures of the image as of now is being experimented by many researchers. These contents leads one to extract the exact feature of the image which can be well utilized to compare with all images available in the database by means of some similarity measures like Euclidean distance, sum of absolute difference etc. The tremendous use of images in the digital world of today has proved the CBIR as very useful in several applications like Finger print recognition, Iris Recognition, face recognition, palm print recognition, speaker identification, pattern matching and recognition etc.

There are various approaches which have been experimented to generate the efficient algorithm for image feature extraction in CBIR. These approaches advocate different ways of extracting features of the images to improve the result in the form of better match of the query image in the large database. Some papers discuss the variation in the similarity measures in order to have its lesser complexity and better match [5-15]. Methods of feature extraction using Vector Quantization [16], bit truncation coding [17,18],Walsh Transform[20,21] has also provided the new horizon to the feature extraction methodology. The method of sectorization has already been experimented on DCT [22], DST [23], DCT-DST Plane [24] , Haar Wavelet [25] and Kekre's Transform [26] earlier.

This paper proposes the use of sectorization of DST Wavelet for feature extraction in CBIR. The outcome of which has been compared with the DST sectorization performance.

II. DST wavelet

A. *Generation of DST Wavelet[4]*

The wavelet analysis procedure is to adopt a wavelet prototype function, called an *analyzing wave* or *mother wave*. Other wavelets are produced by translation and contraction of the mother wave. By contraction and translation infinite set of functions can be generated. This set of functions must be orthogonal and this condition qualifies a transform to be a wavelet transform. Thus there are only few functions which satisfy this condition of orthogonality. Generation of DST Wavelet transform matrix of size N^2xN^2 from DST matrix of size NxN is given in [4]. However in this case we require DST matrix of size 128x128 where 128 is not a square. To resolve this situation, this paper proposes an algorithm to generate discrete sine wavelet transform from discrete sine transform of 8x8 and 16x16.

In this paper the DST Wavelet has been generated using the contraction and translation. Due to the size of images in the database is 128x128 we need the wavelet transform to be of size 128x128. The 128x128 Wavelet transform matrix generated from 16x16 orthogonal DST matrix and 8x8 DST matrix. First 16 rows of Wavelet transform matrix are generated by repeating every column of DST matrix of dimension 16x16, 8 times. To generate next 17 to 32 rows, second row of DST (8X8) is translated using groups of 8 columns with horizontal and downward shifts. To generate next 33 to 48 rows, third row of DST (8X8) matrix is used in the same manner. Like wise to generate last 113 to 128 rows, 8th row of transform DST (8x8) matrix is used. Note that by repeating every column of the basic transform 8 times we get global components. Other wavelets are generated by using rows of DST matrix of size 8x8 giving local components of the DST Wavelet.

III. Feature Vector Generation

The figure given below shows the formal steps of feature vector generation in brief.

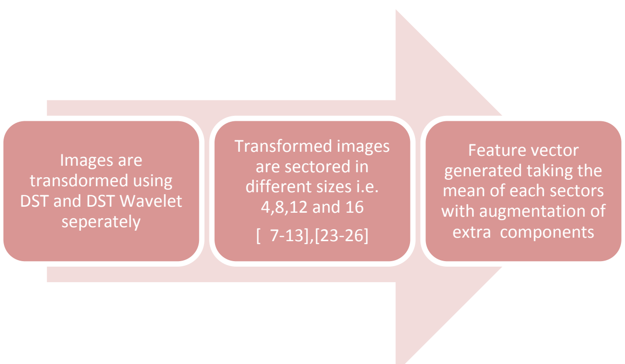


Figure 1 Steps to generate the feature vectors

A. *Sectorization [7-13][23-28]*

The individual components of row and column wise transformed (DST and DST Wavelet separately) images are distributed into different co-ordinates of Cartesian coordinate system according to their sign change to form four sectors. The even rows/columns components of the transformed image and the odd rows/columns components of the transformed images are checked for positive and negative signs. The even and odd DST values are assigned to each quadrant

The division of each of these 4 sectors into 2 partitions forms the 8 sectors which distributes the transformed image components into its appropriate sectors.

Continuing the same division concept further the 4 sectors already generated has been divided into 3 parts with consideration of 30 degree angle to generate sector sizes of 12.

Each sector of 8 sectors are individually divided into two to obtain 16 sectors. For each individual sector sizes the mean of each sectors are taken as the feature vector component. The feature vector components of each plane i.e. R, G and B has been calculated and concatenated together with the average of first row/column and last row/column to form the final feature vector for each method separately. The size of the feature vector and its component varies for each sector sizes the maximum sector size will have maximum feature vector components.

The feature database consists of feature vectors of all images. The features of the query image extracted are compared with the feature database using similarity measures. There are two similarity measures used in this experiment i.e. Eucledian distance (ED) and sum of absolute difference (AD) as given in the equation (1) and (2) shown below:

$$ED(P,Q) = \sum_{i=1}^{N} (Pi - Qi)^2 \qquad (1)$$

$$AD(P,Q) = \sum_{i=1}^{N} |Pi - Qi| \qquad (2)$$

The match of query image feature with the feature database with the minimum value of ED/AD gives the perfect match.

The performance measure of the algorithms proposed are done with the calculation of the precision-recall and LIRS (Length of initial relevant string of retrieval) and LSRR (Length of string to recover all relevant images in the database) refer equations (3) to (6).

$$Precision = \frac{\text{Number of relevant images retrieved}}{\text{Total Number of images retrieved}} \qquad (3)$$

$$Recall = \frac{\text{Number of relevant images retrieved}}{\text{Total number of relevant images in database}} \qquad (4)$$

$$LIRS = \frac{\text{Length of initial relevant string of images}}{\text{Total relevant images in the Database}} \qquad (5)$$

$$LSRR = \frac{\text{Length of string to recover all relevant images}}{\text{Total images in the Database}} \qquad (6)$$

The performance of the proposed methods are checked by means of calculating the class wise average performance and overall average performance of each approach with respect to the transformation method applied, the way of applying the transformation i.e. row wise, column wise, full, sector sizes used and type of similarity measures used.

IV. Experimental Results

A. *Image Database*

The sample Images of the augmented Wang database [29] consists of 1055 images having 12 different classes such as Cartoons, Flowers, Elephants, Barbie, Mountains, Horses, Buses, Sunset, Tribal, Beaches, Monuments and Dinosaur shown in the Figure 2.

Figure 2. Sample images in the Database

The class wise distribution of all images in its respective classes such as ther are 46 images of cartoon in the class, there are 100 images for flower and so on are shown in the Figure 3 below.

Cartoon 46	Flower 100	Elephants 100	Barbie 59	Mountains 100	Horse 100	Bus 100
	Sunset 51	Tribal 100	Beaches 99	Monuments 100	Dinosaur 100	

Figure3. Class wise distribution of images

B. Sectorization of Row wise Transformed (DST) images.

The class wise precision-recall cross over point plot for sectorization of row wise DST transformed images has been shown in the Figure 4. The x axis of the plot denotes the class of images in the database and the y axis denotes the average precision-recall cross over point plots of five randomly selected images per class. These values are taken into percentage. Comparing all classes of images it is found that the dinosaur class has the best retrieval result of more than 80% for 12 and 16 sectors with sum of absolute difference as similarity measure. Other classes like flower, horse, sunset and elephants have the retrieval up to 68% (12 and 16 sectors with AD), 55% (4,8,12 sectors with ED), more than 50% (16 sectors with AD) and more than 50% (12 sectors with AD) respectively. Looking at the LIRS plot for the same (as shown in the Figure 5) to check the performances it has been found that the initial length of string containing the relevant images must be more. In this case the higher value of LIRS is very clearly visible for the best performer class i.e. dinosaur class. The LIRS varies for all other classes from 1% to 20% which indicates that the first image is always relevant image. The Figure 6 shows the LSRR Plot checks for the length of string containing all relevant images retrieved. It must be minimum which is very much seen for dinosaur class.

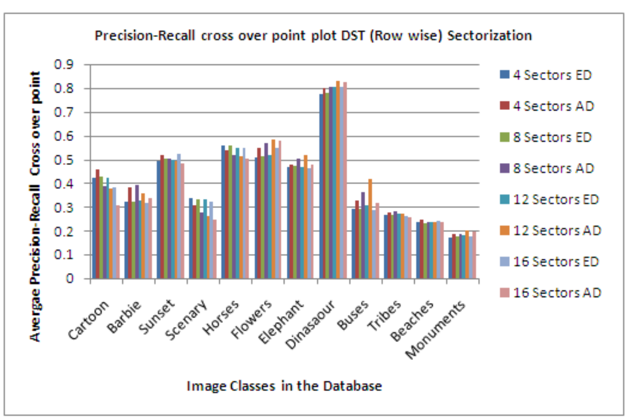


Figure4. Class wise Average Precision-Recall cross over point plot of DST row wise sectorization for all sector sizes with respect to similarity measures i.e. Euclidian distance (ED) and sum of absolute difference (AD)

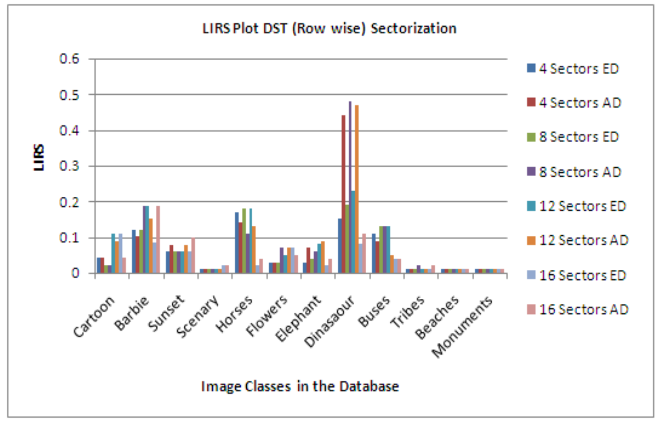


Figure5. Class wise LIRS plot of DST row wise sectorization for all sector sizes with respect to similarity measures i.e. Euclidian distance (ED) and sum of absolute difference (AD)

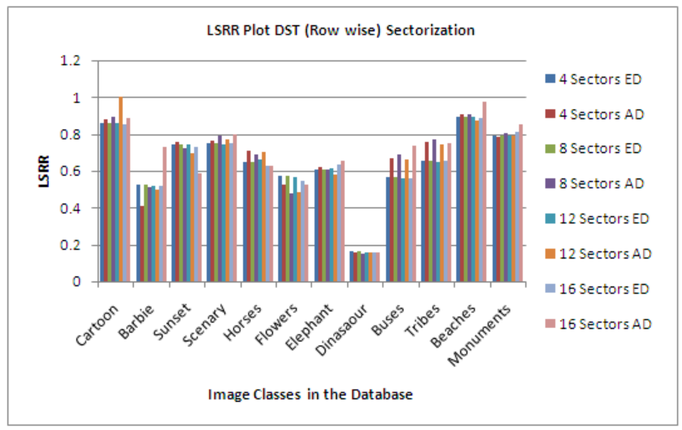


Figure 6. Class wise LSRR plot of DST row wise sectorization for all sector sizes with respect to similarity measures i.e. Euclidian distance (ED) and sum of absolute difference (AD)

C. Sectorization of Row wise Transformed (DST Wavelet) images

This section of the paper discusses the performance of the DST Wavelet (row wise) sectorization with respect to all three performance measuring parameters namely average precision-recall cross over point plot (see Figure7), LIRS(see Figure8) and LSRR(see Figure9). Comparing the performance of all the classes within once again the dinosaur class outperforms all classes with 90% retrieval for 12 sectors with sum of absolute difference as similarity measure. This performance is better than the simple DST (row wise) sectorization as discussed in section 4.2. The flower class has retrieval of more than 65% in DST Wavelet whereas it is below 60% in the case of DST. Horse class has the result of 50%,The cartoon class has improved a lot goes up to 60% compared to only 45% in normal DST sectorization. The LIRS and LSRR shows its relevant plots in the Figure8 and Figure9.

D. Sectorization of Column wise Transformation (DST)

There are 12 classes of the images used in the database. The performance of the algorithm varies from class to class. The class wise average precision-recall cross over point plotted din the Figure10 shows the performance of the sectorization of DST (column wise) in various sectors.

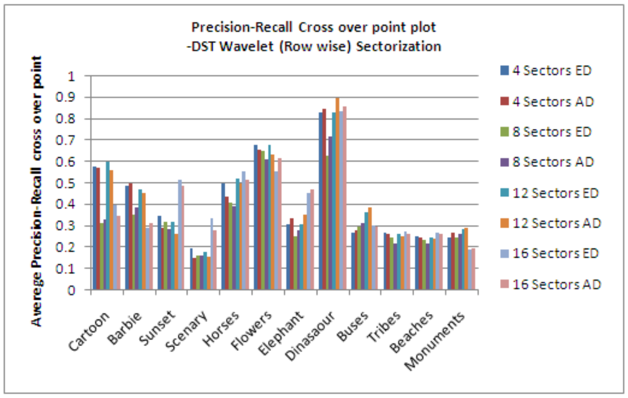


Figure7. Class wise Average Precision-Recall cross over point plot of DST Wavelet row wise sectorization for all sector sizes with respect to similarity measures i.e. Euclidian distance (ED) and sum of absolute difference (AD)

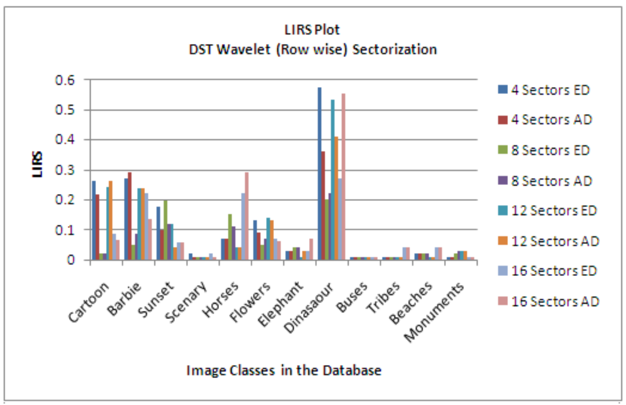


Figure 8. Class wise LIRS plot of DST Wavelet row wise sectorization for all sector sizes with respect to similarity measures i.e. Euclidian distance (ED) and sum of absolute difference (AD)

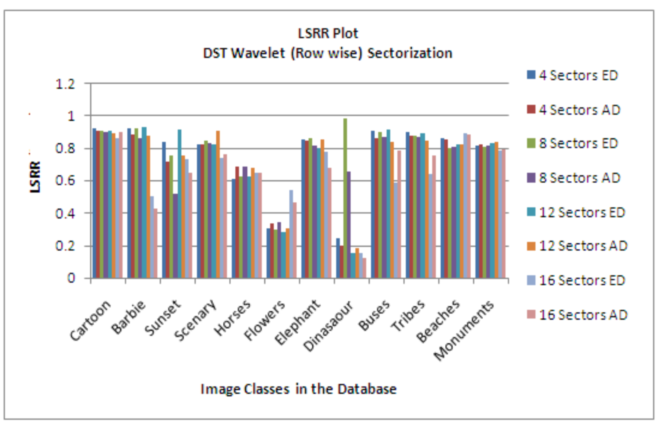


Figure9. Class wise LSRR plot of DST Wavelet row wise sectorization for all sector sizes with respect to similarity measures i.e. Eucledian distance (ED) and sum of absolute difference (AD)

Looking at the class wise performance the Dinasaour class has the best retrieval close to 80%, Flower class reaches up to 70% (16 sectors with AD), sunset and horses class has the retrieval more than 50%.

The elephant class has the resultant retrieval rate close to 50% whereas the performance for the Barbie class is more than 40%. The LIRS and LSRR which keeps check on the performance evaluation of the method shown in the Figure11 and Figure 12.The maximum f LIRS has been achieved for the dinosaur class for the combination of all sectors and the sum of absolute difference.

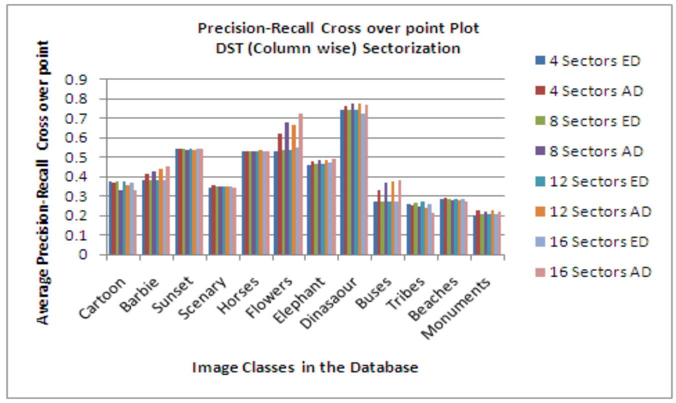


Figure10. Class wise Average Precision-Recall cross over point plot of DST column wise sectorization for all sector sizes with respect to similarity measures i.e. Euclidian distance (ED) and sum of absolute difference (AD)

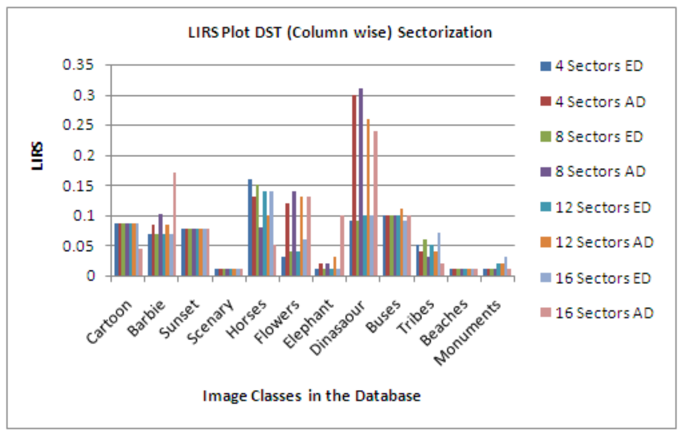


Figure 11. Class wise LIRS plot of DST column wise sectorization for all sector sizes with respect to similarity measures i.e. Euclidian distance (ED) and sum of absolute difference (AD)

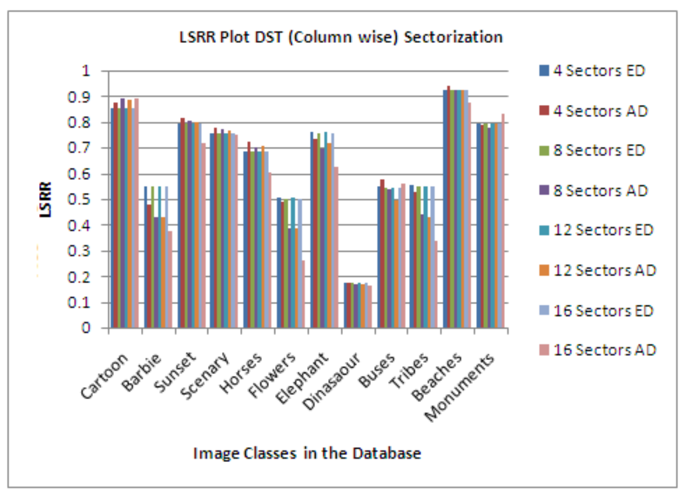


Figure12. Class wise LSRR plot of DST column wise sectorization for all sector sizes with respect to similarity measures i.e. Euclidian distance (ED) and sum of absolute difference (AD)

E. Sectorization of Column wise Transformation (DST Wavelet).

The sectorization of DST Wavelet has better performance of average precision-recall cross over points as it can be easily seen in the Figure 13.There is increase in the performance of the Dinosaur, Flowers, Cartoon, Barbie classes i.e. 85%(12 and 16 sector with AD), 75% (16 sectors with ED), 42%(16 sectors with AD), 50%(16 sectors with AD) respectively.

The LIRS performances of the method are very interestingly depicts the high rises for the better performance of precision-recall. The minimum value of LSRR for the dinosaur, flower, sunset classes depicts the good performance.

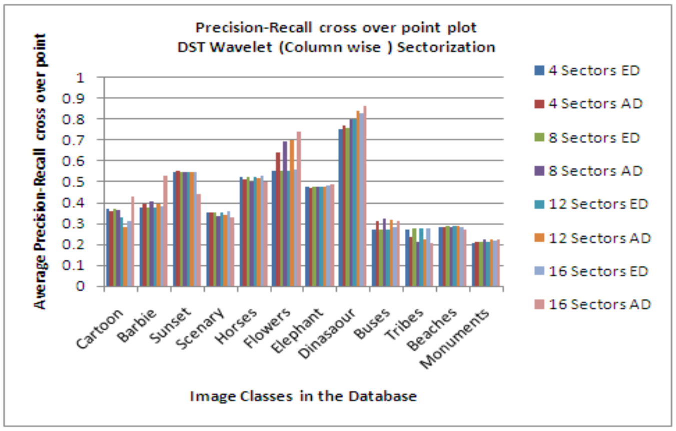


Figure13. Class wise Average Precision-Recall cross over point plot of DST Wavelet column wise sectorization for all sector sizes with respect to similarity measures i.e. Euclidian distance (ED) and sum of absolute difference (AD)

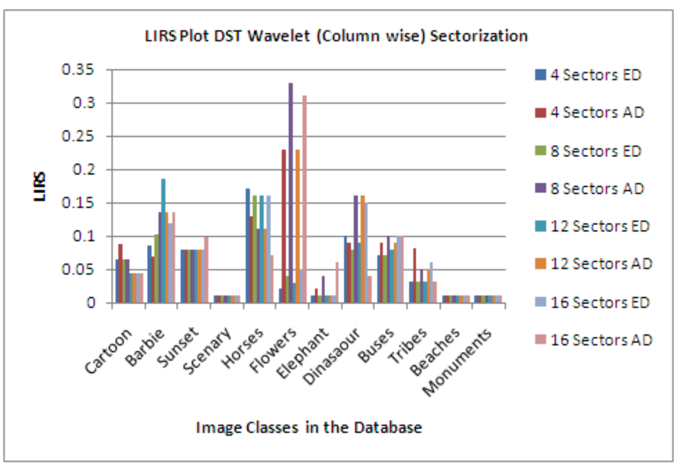


Figure14. Class wise LIRS plot of DST Wavelet column wise sectorization for all sector sizes with respect to similarity measures i.e. Euclidian distance (ED) and sum of absolute difference (AD)

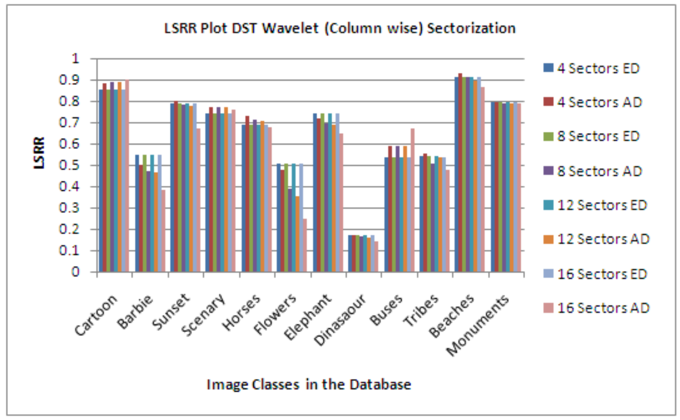


Figure15. Class wise LSRR plot of DST Wavelet column wise sectorization for all sector sizes with respect to similarity measures i.e. Euclidian distance (ED) and sum of absolute difference (AD)

F. Overall Comparison of all Approaches.

The overall performance of all approaches gives very clear idea about the overall average performance of the retrieval rates as shown in the Figure16 – Figure 18. The overall average precision-recall cross over point plot as shown in the Figure16 depicts that on average performance of the retrieval for all methods proposed is 40%. It is observed that In most of the cases the sectorization with the sum of absolute difference as similarity measure has better retrieval than Euclidian distance. As far as sector sizes are concerned all have good performance except for 8 sectors in DST-WT (Row).

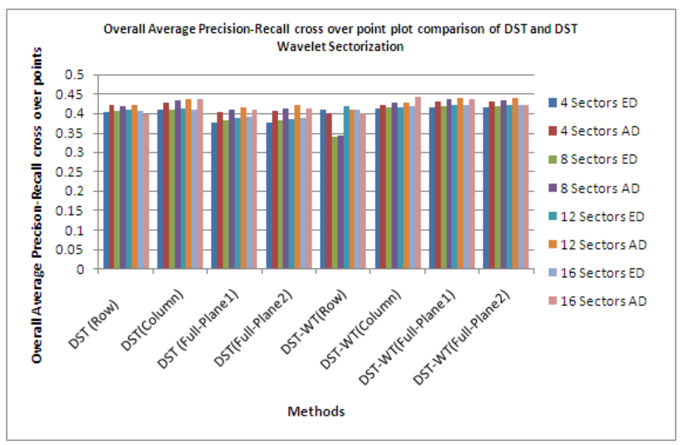


Figure16. Overall Average Precision-Recall cross over point plot comparison of DST and DST Wavelet Sectorization for all sector sizes with respect to similarity measures i.e. Euclidian distance (ED) and sum of absolute difference (AD)

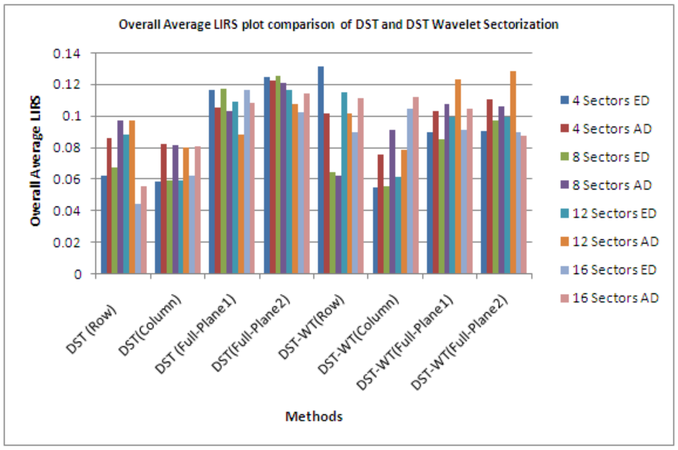


Figure17. Overall Average LIRS plot comparison of DST and DST Wavelet Sectorization for all sector sizes with respect to similarity measures i.e. Euclidian distance (ED) and sum of absolute difference (AD)

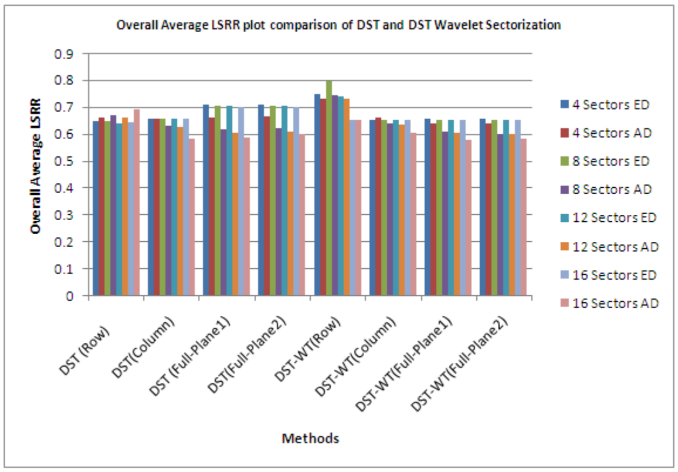


Figure18. Overall Average LSRR plot comparison of DST and DST Wavelet Sectorization for all sector sizes with respect to similarity measures i.e. Euclidian distance (ED) and sum of absolute difference (AD)

V. CONCLUSION

This paper discusses the new idea of generating the DST wavelet transform and sectoring the transformed image into 4,8,12 and 16 sectors. The results obtained clearly show that the performance of DST Wavelet transform is better than DST using sectorization of full-Plane1 and full-plane2 transformations with the sum of absolute difference giving best results close to 45% (see the Figure 16). The similarity measure plays vital role in the applications of CBIR. Similarity measures provide the efficient and faster access and matching of the images in the database. We have employed and analyzed two similarity measures i.e Euclidian Distance (ED) and sum of absolute difference (AD).The sum of absolute difference as shown in the equation (2) has lesser complexity of calculation and better retrieval rates in the approaches applied for all sector sizes (see figure 16). The performance measurement of all approaches has been checked by means of LIRS and LSRR which are very powerful tools to comment on the retrieval rate of any algorithm. The maximum LIRS and the Minimum LSRR are the best measures for it.

REFERENCES

[1] Kato, T., "Database architecture for content based image retrieval in Image Storage and Retrieval Systems" (Jambardino A and Niblack W eds),*Proc SPIE 2185*, pp 112-123, 1992.

[2] Ritendra Datta,Dhiraj Joshi,Jia Li and James Z. Wang, " Image retrieval:Idea,influences and trends of the new age",ACM Computing survey,Vol 40,No.2,Article 5,April 2008.

[3] John Berry and David A. Stoney "The history and development of fingerprinting," in *Advances in Fingerprint Technology*, Henry C. Lee and R. E. Gaensslen, Eds., pp. 1-40. CRC Press Florida, 2nd edition, 2001.

[4] H.B.Kekre, Archana Athawale, Dipali sadavarti, "Algorithm to generate wavelet transform from an orthogonal transform", International journal of Image processing, Vol.4(4), pp.444-455

[5] H. B. Kekre, Dhirendra Mishra, "Digital Image Search & Retrieval using FFT Sectors" published in proceedings of National/Asia pacific conference on Information communication and technology(NCICT 10) 5TH & 6TH March 2010.SVKM'S NMIMS MUMBAI

[6] H.B.Kekre, Dhirendra Mishra, "Content Based Image Retrieval using Weighted Hamming Distance Image hash Value" published in the proceedings of international conference on contours of computing technology pp. 305-309 (Thinkquest2010) 13th & 14th March 2010.

[7] H.B.Kekre, Dhirendra Mishra,"Digital Image Search & Retrieval using FFT Sectors of Color Images" published in International Journal of Computer Science and Engineering (IJCSE) Vol. 02,No.02,2010,pp.368-372 ISSN 0975-3397 available online at http://www.enggjournals.com/ijcse/doc/IJCSE10-02- 02-46.pdf

[8] H.B.Kekre, Dhirendra Mishra, "CBIR using upper six FFT Sectors of Color Images for feature vector generation" published in International Journal of Engineering and Technology(IJET) Vol. 02, No. 02, 2010, 49-54 ISSN 0975-4024 available online at http://www.enggjournals.com/ijet/doc/IJET10-02- 02-06.pdf

[9] H.B.Kekre, Dhirendra Mishra, "Four walsh transform sectors feature vectors for image retrieval from image databases", published in international journal of computer science and information technologies (IJCSIT) Vol. 1 (2) 2010, 33-37 ISSN 0975-9646 available online at http://www.ijcsit.com/docs/vol1issue2/ijcsit2010010201.pdf

[10] H.B.Kekre, Dhirendra Mishra, "Performance comparison of four, eight and twelve Walsh transform sectors feature vectors for image retrieval from image databases", published in international journal of Engineering, science and technology(IJEST) Vol.2(5) 2010, 1370-1374 ISSN 0975-5462 available online at http://www.ijest.info/docs/IJEST10-02-05-62.pdf

[11] H.B.Kekre, Dhirendra Mishra, "density distribution in walsh transfom sectors ass feature vectors for image retrieval", published in international journal of compute applications (IJCA) Vol.4(6) 2010, 30-36 ISSN 0975-8887 available online at http://www.ijcaonline.org/archives/volume4/number6/829-1072

[12] H.B.Kekre, Dhirendra Mishra, "Performance comparison of density distribution and sector mean in Walsh transform sectors as feature vectors for image retrieval", published in international journal of Image Processing (IJIP) Vol.4(3) 2010, ISSN 1985-2304 available online at http://www.cscjournals.org/csc/manuscript/Journals/IJIP/Volume4/Issue 3/IJIP-193.pdf

[13] H.B.Kekre, Dhirendra Mishra, "Density distribution and sector mean with zero-sal and highest-cal components in Walsh transform sectors as feature vectors for image retrieval", published in international journal of Computer scienece and information security (IJCSIS) Vol.8(4) 2010, ISSN 1947-5500 available online http://sites.google.com/site/ijcsis/vol-8-no-4-jul-2010

[14] Arun Ross, Anil Jain, James Reisman, "A hybrid fingerprint matcher," *Int'l conference on Pattern Recognition (ICPR)*, Aug 2002.

[15] A. M. Bazen, G. T. B.Verwaaijen, S. H. Gerez, L. P. J. Veelenturf, and B. J. van der Zwaag, "A correlation-based fingerprint verification system," *Proceedings of the ProRISC2000 Workshop on Circuits, Systems and Signal Processing*, Veldhoven, Netherlands, Nov 2000.

[16] H.B.Kekre, Tanuja K. Sarode, Sudeep D. Thepade, "Image Retrieval using Color-Texture Features from DCT on VQ Codevectors obtained by Kekre's Fast Codebook Generation", ICGST International

Journal on Graphics, Vision and Image Processing (GVIP), Available online at http://www.icgst.com/gvip

[17] H.B.Kekre, Sudeep D. Thepade, "Using YUV Color Space to Hoist the Performance of Block Truncation Coding for Image Retrieval", IEEE International Advanced Computing Conference 2009 (IACC'09), Thapar University, Patiala, INDIA, 6-7 March 2009.

[18] H.B.Kekre, Sudeep D. Thepade, "Image Retrieval using Augmented Block Truncation Coding Techniques", ACM International Conference on Advances in Computing, Communication and Control (ICAC3-2009), pp.: 384-390, 23-24 Jan 2009, Fr. Conceicao Rodrigous College of Engg., Mumbai. Available online at ACM portal.

[19] H.B.Kekre, Tanuja K. Sarode, Sudeep D. Thepade, "DCT Applied to Column mean and Row Mean Vectors of Image for Fingerprint Identification", International Conference on Computer Networks and Security, ICCNS-2008, 27-28 Sept 2008, Vishwakarma Institute of Technology, Pune.

[20] H.B.Kekre, Sudeep Thepade, Archana Athawale, Anant Shah, Prathmesh Velekar, Suraj Shirke, " Walsh transform over row mean column mean using image fragmentation and energy compaction for image retrieval", International journal of computer science and engineering (IJCSE),Vol.2.No.1,S2010,47-54.

[21] H.B.Kekre, Vinayak Bharadi, "Walsh Coefficients of the Horizontal & Vertical Pixel Distribution of Signature Template", In Proc. of Int. Conference ICIP-07, Bangalore University, Bangalore. 10-12 Aug 2007.

[22] H.B.Kekre, Dhirendra Mishra, "DCT Sectorization for feature vector generation in CBIR", International journal of computer Applications (IJCA),Vol.9,No.1,pp.19-26

[23] H.B.Kekre, Dhirendra Mishra, "DST Sectorization for Feature vector generation", Universal journal of computer science and and Engineering Technology (UniCSE),Vol.1, No.1, Oct.2010,pp.6-15,Available online at
http://www.unicse.org/index.php?option=com_content&view=article&id=54&Itemid=27

[24] H.B.Kekre, Dhirendra Mishra, "DCT-DST Plane sectorization of row wise transformed color images in CBIR", International journal of engineering science and technology, Vol.2, No.12, Dec.2010, pp.7234-7244, ISSN No.0975-5462. Available at http://www.ijest.info/docs/IJEST10-02-12-143.pdf

[25] H.B.Kekre, Dhirendra Mishra, "Sectorization of Haar and Kekre's Wavelet for feature extraction of color images in image retrieval", International journal of computer science and information security (IJCSIS), USA, Vol.9, No.2, Feb 2011, pp.180-188, http://sites.google.com/site/ijcsis/volume-9-no-2-feb-2011

[26] H.B.Kekre, Dhirendra Mishra, "Sectorization of Kekre's transform for image retrieval in content based image retrieval", Journal of Telecommunication (JOT),UK, Vol.8, No.1 April 2011, pp. 26-33. http://sites.google.com/site/journaloftelecommunications/volume-8-issue-1-april-2011

[27] H.B.Kekre, Dhirendra Mishra, "Sectorization of DCT- DST Plane for column wise transformed color images in CBIR", International Conference of Technology Systems & Management (ICTSM-2011) held at SVKM's NMIMS Mumbai India, published in Springer Link CCIS 145, pp. 55–60, 2011. Available online at
http://www.springerlink.com/content/m573256n53r07733/

[28] H.B.Kekre, Dhirendra Mishra, "Full DCT sectorization for Feature vector generation in CBIR", Journal of graphics, vision, image processing, Vol.11, No.2, April 2011, pp. 19 – 30 http://www.icgst.com/gvip/Volume11/Issue2/P1151041315.html

[29] Jia Li, James Z. Wang, ``Automatic linguistic indexing of pictures by a statistical modeling approach," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, no. 9, pp. 1075-1088, 2003

AUTHORS PROFILE

H. B. Kekre has received B.E. (Hons.) in Telecomm. Engg. from Jabalpur University in 1958, M.Tech (Industrial Electronics) from IIT Bombay in 1960, M.S.Engg. (Electrical Engg.) from University of Ottawa in 1965 and Ph.D.(System Identification) from IIT Bombay in 1970. He has worked Over 35 years as Faculty and H.O.D. Computer science and Engg. At IIT Bombay. From last 13 years working as a professor in Dept. of Computer Engg. at Thadomal Shahani Engg. College, Mumbai. He is currently senior Professor working with Mukesh Patel School of Technology Management and Engineering, SVKM's NMIMS University vile parle west Mumbai. He has guided 17 PhD.s 150 M.E./M.Tech Projects and several B.E./B.Tech Projects. His areas of interest are Digital signal processing, Image Processing and computer networking. He has more than 350 papers in National/International Conferences/Journals to his credit. Recently twelve students working under his guidance have received the best paper awards. Two research scholars working under his guidance have been awarded Ph. D. degree by NMIMS University. Currently he is guiding 10 PhD. Students. He is life member of ISTE and Fellow of IETE.

Dhirendra Mishra has received his BE (Computer Engg) degree from University of Mumbai. He completed his M.E. (Computer Engg) from Thadomal shahani Engg. College, Mumbai, University of Mumbai. He is PhD Research Scholar and working as Associate Professor in Computer Engineering department of, Mukesh Patel School of Technology Management and Engineering, SVKM's NMIMS University, Mumbai, INDIA. He is life member of Indian Society of Technical education (ISTE), Member of International association of computer science and information technology (IACSIT), Singapore, Member of International association of Engineers (IAENG). He has more than 30 papers in National/International conferences and journals to his credit. His areas of interests are Image Processing, Operating system, Information Storage and Management.

Performance Analysis of UMTS Cellular Network using Sectorization Based on Capacity and Coverage

A.K.M Fazlul Haque, Mir Mohammad Abu Kyum, Md. Baitul Al Sadi, Mrinal Kar
Department of Electronics and Telecommunication Engineering.
Daffodil International University

Md. Fokhray Hossain
Department of Computer Science and Engineering.
Daffodil International University

***Abstract*—Universal Mobile Telecommunications System *(*UMTS) is one of the standards in 3rd generation partnership project (3GPP). Different data rates are offered by UMTS for voice, video conference and other services. This paper presents the performance of UMTS cellular network using sectorization for capacity and coverage. The major contribution is to see the impact of sectorization on capacity and cell coverage in 3G UMTS cellular network. Coverage and capacity are vitally important issues in UMTS cellular Network. Capacity depends on different parameters such as sectorization, energy per bit noise spectral density ratio, voice activity, inter-cell interference and intra-cell interference, soft handoff gain factor, etc and coverage depends on frequency, chip rate, bit rate, mobile maximum power, MS Antenna Gain, EIRP, interference Margin, Noise figure etc. Different parameters that influence the capacity and coverage of UMTS cellular network are simulated using MATLAB 2009a. In this paper, the outputs of simulation for increasing amount of sectorization showed that the number of users gradually increased. The coverage area also gradually increased.**

Keywords-UMTS; Capacit; Coverage and data rates; sectoring.

I. INTRODUCTION

A cellular cell can be divided into number of geographic areas, called sectors. It may be 3 sectors, 4 sectors, 6 sectors etc. When sectorization is done in a cell, interference is significantly reduced resulting in better performance for cellular network. Capacity in WCDMA standards of UMTS refers to maximum number of users per cell, where the area covered by RF signal from Node B or UE (User Equipment) is called coverage area of UMTS. Capacity and coverage are two dynamic phenomena in UMTS network. Parameters that define capacity and coverage of UMTS are dynamic in nature, where increasing or decreasing values of these parameters affects capacity and coverage of UMTS cellular network. One of the parameters is sectorization in UMTS. There are some works on sectorization scheme [1-4].

Bo Hagerman, Davide Imbeni and Jozsef Barta considered WCDMA 6-sector deployment case study of a real installed UMTS-FDD network [1]. Romeo Giuliano, Franco Mazzenga, Francesco Vatalaro described Adaptive Cell Sectorization for UMTS Third Generation CDMA Systems [2]. Achim Wacker, Jaana Laiho-Steffens, Kari Sipila, and Kari Heiska considered the impact of the base station sectorisation on WCDMA radio network performance [3]. S. Sharma, A.G. Spilling and A.R. Nix considered Adaptive Coverage for UMTS Macro cells based on Situation Awareness [4]. Most of the works analyzed the performance considering sectors with static parameters but it is needed to analyze the performance along with all dynamic parameters.

This paper optimizes the performance of both capacity and coverage of UMTS not only considering sectors but also with dynamic parameters as energy per bit noise spectral density ratio, voice activity, inter-cell interference, soft handoff factor, and data rates.

II. BACKGROUND

A. Capacity in WCDMA for UMTS:

As the downlink capacity of UMTS is related to transmit power of Node B and uplink capacity is related to numbers of users, uplink capacity is considered in this paper.

If the number of users is N_s then for a single CDMA cell, the number of users will be [5],

$$N_s = 1 + \left(\frac{W/R}{E_b/N_o} - \frac{\eta}{S}\right)\frac{1}{\alpha} \qquad \text{(i)}$$

Where, $N_{s=}$total number of users, W=chip rate,

R= base band information bit rate, E_b/N_o=Energy per bit to noise power spectral density ratio, η= background thermal noise, S=signal power=S1-P (d)-shadow fading, S1=UE power, P (d) =Propagation loss.

For WCDMA, the chip rate is 3.84 Mcps [8], and the channel bandwidth is 5 MHz [8]. It is also necessary to consider the affects of multiple cells or intra-cell interference (β)[12], cell sectoring(D)[6], soft handover factor(H)[11], Array antenna gain (A_g)[10].Thus the capacity for WCDMA in UMTS yields:

$$N_s = 1 + \left(\frac{W/R}{E_b/N_o} - \frac{\eta}{S}\right)\frac{1}{(1+\beta)\alpha} \times D \times H \times A_g \quad \text{(ii)}$$

B. Coverage and data rates in WCDMA for UMTS:

UMTS offered different data rates for multi services. Table 1 shows different standard bit rates offered by UMTS. Higher class of service makes cell radius small resulting in small coverage area. If different class of services is classified in terms of coverage area, it will look like figure 1.

TABLE 1: DIFFERENT CLASSES OF SERVICES

Bit Rate(Kbps)	Class
12.2	Class 5
32	Class 4
64	Class 3
144	Class 2
384	Class 1

Figure 1 shows that for service class 1, maximum distance is observed by UE (User Equipment) from Node B. Similarly for service class 2 and service class 3, UEs maintain maximum distance from Node B. From this figure it is clear that different coverage areas are needed to maintain different data rates. So coverage area needs to increase for better class of services. This paper optimizes the coverage area for particular services with sectors.

Figure 2 shows a UMTS cell where Node B received power (P_R) from User Equipment (UE). The Node B sensitivity is the power level for minimum signal necessary at the input of the Node B receiver to meet requirements in terms of E_b/N_o, processing gain (G_p) and Node B interference and noise power given as [8]

Node B sensitivity=$E_b/N_o - G_p + N_{\text{Node B interference and noise power}}$

Where G_p=Processing gain

$$= 10\log\left(\frac{Chiprate}{Bitrate}\right) = 10\log\left(\frac{3.84Mcps}{R}\right)$$

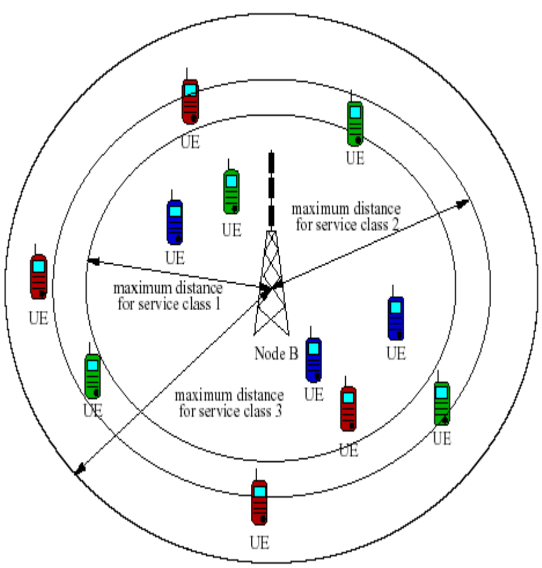


Figure 1: Different Classes of Services vs. Maximum Distance

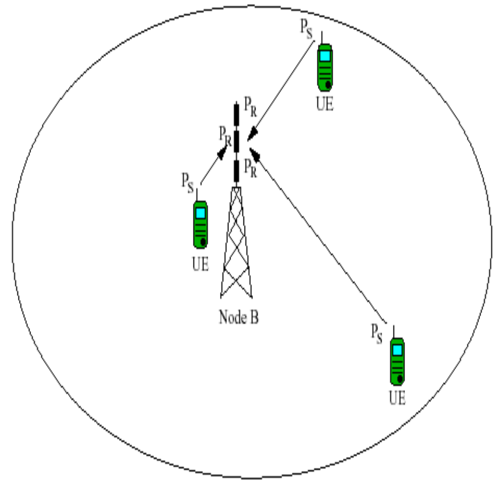


Figure 2: UMTS cell

Now the maximum allowable path loss for Node B,

L_P =EIRP-Node b sensitivity + G_p - fast fading margin----- (iii)

from radio propagation model, Path loss for dense urban area [5],

$$L = 46.3 + 33.9\log(f_c) - 13.82\log h_b - 3.2 [\log(11.75 h_{UE})]^2 + 4.97 + (44.9 - 6.55 \log h_{NodeB})\log d + 3$$

-- (iv)

From equation (iii) and (iv) a relationship can be expressed for coverage and data rates in dense urban case,

$$46.3 + 33.9\log(f_c) - 13.82\log h_b - 3.2 [\log(11.75 h_{UE})]^2 + 4.97 + (44.9 - 6.55 \log h_{NodeB})\log d + 3 = \text{EIRP - NodeB sensitivity} + 10\log\left(\frac{\text{Chip rate}}{R}\right) \text{- Fast fading Margin}$$

where d is the coverage radius and R is the data rates.

After calculating the cell range d, the coverage area can be calculated. The coverage area for one cell in hexagonal configuration can be estimated with [9]

Coverage area, S =K.d^2

where S is the coverage area, d is the maximum cell range, and K is a constant. In Table 2, some of the K values are listed.

TABLE 2: K VALUES FOR THE SITE AREA CALCULATION [9]:

Site configuration	Omni or no sector	Two sectors	Three sectors	Four sectors
Value of K	2.6	1.3	1.95	2.6

III. SIMULATIONS AND RESULTS

The analysis has been done for capacity and coverage with sectoring cell for dense urban using MATLAB R2009a. The simulated values for sectorization are shown in Table 3, Table 4, Table 5, Table 6 and Table 7. The performances are also described in Figure 3, Figure 4, Figure 5, Figure 6, and Figure 7. The algorithms of the evaluation process have also been introduced in appendix.

Figure 3 shows that Energy per bit to noise spectral density ratio (E_b/N_o) needs to maintain small value for increasing number of simultaneous 384 Kbps users. From this figure it is observed that for dynamic value of E_b/N_o with changing the sectors, the number of simultaneous 384 Kbps data users increased or decreased. For example, if E_b/N_o value is 4 db, then for 6 sectors, the number of simultaneous users will be 88 but for 3 sectors, the number of simultaneous users will be 45. Thus the dynamic values of E_b/N_o can be the increasing or decreasing factors in UMTS and sectorization scheme can be effective in this case.

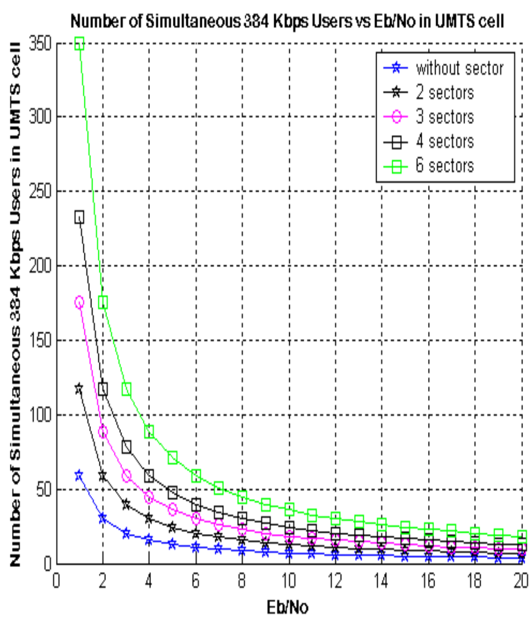


Figure 3: Number of simultaneous 384 Kbps users vs. Eb/No in sectors cell

TABLE 3: SIMULATED VALUES FOR NUMBER OF SIMULTANEOUS 384 KBPS USERS VS. EB/NO IN SECTORS CELL

Energy per bit to Noise spectral density ratio(E_b/N_o)	Users without sector	Users with 2 sectors	Users with 3 sectors	Users with 4 sectors	Users with 6 sectors
1	59.065	117.13	175.19	233.26	349.39
4	15.516	30.032	44.548	59.065	88.097
8	8.2581	15.516	22.774	30.032	44.548
10	6.8065	12.613	18.419	24.226	35.839
14	5.1475	9.2949	13.442	17.59	25.885
16	4.629	8.2581	11.887	15.516	22.774
18	4.4156	7.4516	10.677	13.903	20.355
20	3.9032	6.8065	9.7097	12.613	18.419

The interference from other cell is known as inter-cell interference (β). For multi-cell configuration, the number of outer cells can reduce cell capacity in UMTS. Figure 4 shows, for increasing demand of users the value of β in UMTS needs to be small. Figure 4 also represents dynamic inter-cell interference with changing of sectors, where number of simultaneous 384 Kbps data users increases or decreases. From Figure 4 it has been observed that for increasing value of β, it is needed to increase sectors.

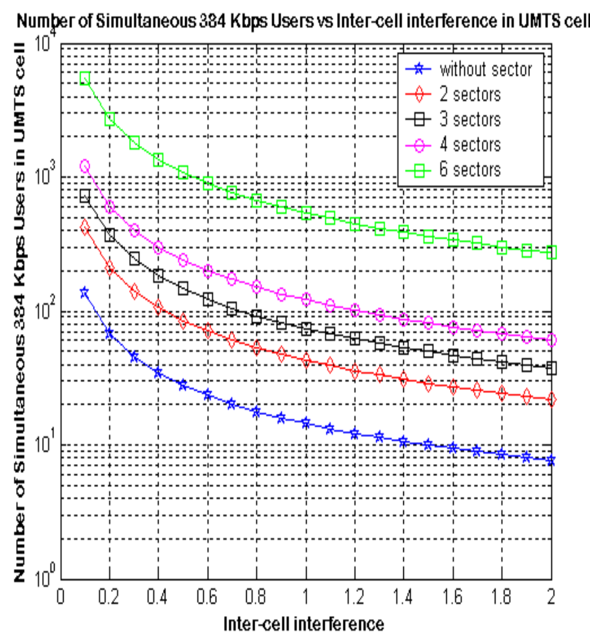


Figure 4: Number of simultaneous 384 Kbps users vs.inter-cell interference in sectors cell

The overlapped cell can lead an extra power thus introducing soft handover factor (H) in a UMTS cell. The value of H in UMTS can be a factor to increase the number of users. Figure 5 shows that for increasing H and changing value of sectorization the number of simultaneous 384 Kbps data users increases.

For example, if H value is 2.5 db, then for 2 sectors the number of simultaneous users will be 195 but for 4 sectors the number of simultaneous users will be 388.

Figure 6 shows that the number of voice users depends on the value voice activity factors (α).This is true only for 12.2 Kbps voice users, not for data users, as for data services it will always be 1.

Figure 3 also shows that, for increasing amount voice users the value of α in UMTS needs to as small as possible. Varying α and changing the sectors the number of simultaneous voice users from figure 6 is observed.

TABLE 4: SIMULATED VALUES FOR NUMBER OF SIMULTANEOUS USERS VS. INTER-CELL INTERFERENCE IN SECTORS CELL

Inter-cell interference	**Users without sector**	**Users with 2 sectors**	**Users with 3 sectors**	**Users with 4 sectors**	**Users with 6 sectors**
0.1	135.33	419.6	730.73	1201	5401
0.5	27.866	84.721	146.95	241	1081
1	14.433	42.86	73.973	121	541
1.5	9.9552	28.907	49.649	81	361
1.7	8.9017	25.624	43.925	71.588	53.941
2.0	7.7164	21.93	37.486	61	271

TABLE 5: SIMULATED VALUES FOR NUMBER OF SIMULTANEOUS USERS VS. SOFT HANDOVER FACTOR IN SECTORS CELL

Soft Hand-over Factor	**Users without sector**	**Users with 2 sectors**	**Users with 3 sectors**	**Users with 4 sectors**	**Users with 6 sectors**
0.1	4.871	8.7419	12.613	16.484	24.226
0.4	16.484	31.968	47.452	62.935	93.903
1	39.71	78.419	117.13	155.84	233.26
1.5	59.065	117.13	175.19	233.26	349.39
2.5	97.774	194.55	291.32	388.1	581.65
3	117.13	233.26	349.39	465.52	697.77

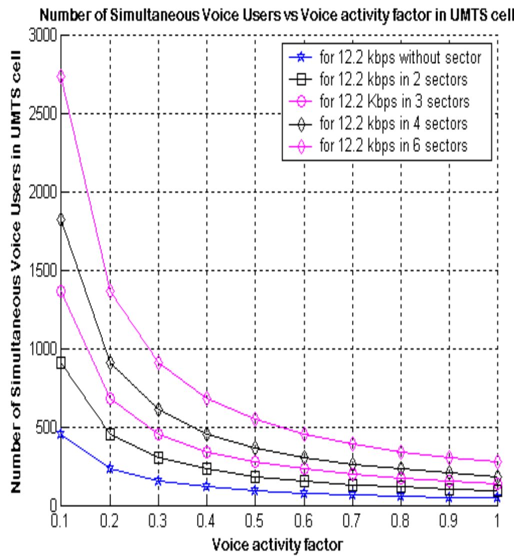


Figure 6: Number of simultaneous voice users vs. voice activity factor in sectors cell.

TABLE 6: SIMULATED VALUES FOR NUMBER OF SIMULTANEOUS USERS VS. VOICE ACTIVITY FACTOR IN SECTORS CELL

Voice activity factor	**Users without sector**	**Users with 2 sectors**	**Users with 3 sectors**	**Users with 4 sectors**	**Users with 6 sectors**
0.2	228.31	455.63	682.94	910.26	1364.9
0.4	114.66	228.31	341.97	455.63	682.94
0.6	76.771	152.54	228.31	304.09	455.63
0.8	57.828	114.66	171.49	228.31	341.97
1	46.463	91.926	137.39	182.85	273.78

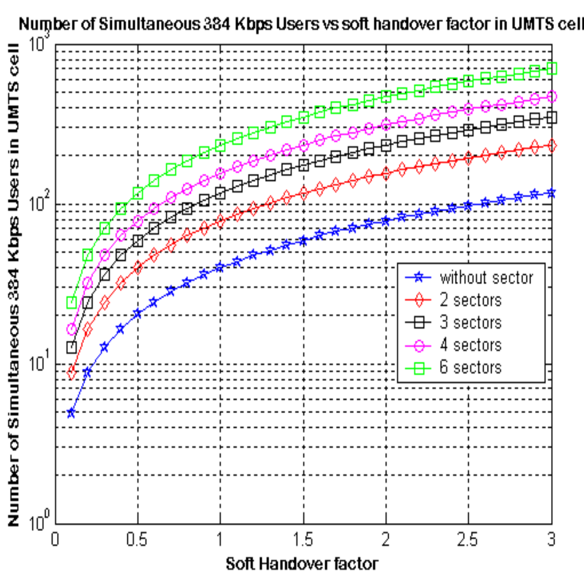


Figure 5: Number of simultaneous 384 Kbps users vs. soft handover factor in sectors cell

TABLE 7: SIMULATED VALUES FOR COVERAGE VS. DATA RATES IN DENSE URBAN USING COST 231 MODEL IN SECTORS CELL

Data rate (Kbps)	Cell range in (meter)	Cell Area without sector (meter^2)	Cell Area with 2 sectors (meter^2)	Cell Area with 3 sectors (meter^2)	Cell Area with 4 sectors (meter^2)
200	773.67	598.57	778.14	1167.2	1556.3
400	635.43	403.77	524.9	78.7.34	1049.8
600	566.31	320.71	416.92	625.38	833.84
800	521.88	272.36	354.07	531.1	708.14
1000	489.83	239.94	311.92	467.88	623.84
1200	465.12	216.33	281.23	421.85	562.47
1400	445.19	198.2	257.66	386.49	515.31
1600	428.63	183.72	238.84	358.26	477.67
1800	414.53	171.83	223.38	335.07	446.76
2000	402.31	161.85	210.41	315.61	420.81

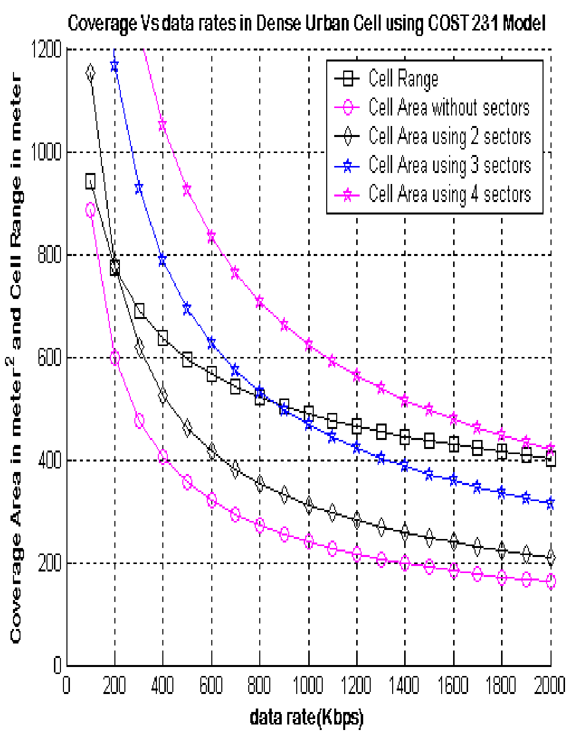


Figure 7: Coverage vs. bit rates for dense urban using COST 231 model in sectors cell

Finally, consider for coverage vs. data rates in dense urban area, where operating frequency is considered 2000 MHz with COST 231 Model as a radio propagation model. In Figure 7 the x axis represents data rate in Kbps and y axis represents coverage area in meter^2 with cell radius in meter. Parameters that are related to coverage setting first, then for increasing data rates in x axis, the coverage area is observed in y axis. From figure 1, it is known that for higher data rates, the coverage will be smaller. It is true only when cell area is considered without sectors. This phenomenon is revealed by figure 7. Figure 7 also shows, for higher data rates comprehensive coverage area is found with increasing sectors.

IV. Conclusion

In this paper, the performance analyses in coverage and capacity of UMTS cellular network using sectorization have been simulated and evaluated for dynamic parameters. The number of simultaneous users increases or decreases for increasing or decreasing sectors with dynamic parameters. Coverage has been estimated for dense urban using COST 231 model where higher data rates need higher processing gain resulting in smaller coverage area. But increasing sectors with same parameters makes extensive coverage for higher data rates.

References

[1] Bo Hagerman, Davide Imbeni and Jozsef Barta "WCDMA 6 - sector Deployment-Case Study of a Real Installed UMTS-FDD Network" IEEE Vehicular Technology Conference, spring 2006, page(s): 703 - 707.

[2] S. Sharma, A.G. Spilling and A.R. Nix "Adaptive Coverage for UMTS Macro cellsbased on Situation Awareness". IEEE Vehicular Technology Conference, spring 2001, page(s):2786 - 2790

[3] A. Wacker, J. Laiho-Steffens, K. Sipila, K. Heiska, "The impact of the base station sectorisation on WCDMA radio network performance", IEEE Vehicular Technology Conference ,September 1999,page(s): 2611 - 2615 vol.5.

[4] Romeo Giuliano, Franco Mazzenga, Francesco Vatalaro, "Adaptive cell sectorization for UMTS third generation CDMA systems" IEEE Vehicular Technology Conference, May 2001, page(s): 219 - 223 vol.1.

[5] T.S.Rappaport, "Wireless Communications Principles and Practice"-Second Edition, Prentice Hall.

[6] Bernard Sklar, "Digital Communications - Fundamentals and Applications"- Second Edition Prentice Hall.

[7] Ingo Fo, Marc Schinnenburg, Bianca Wouters "Performance evaluation of soft handover in a realistic UMTS network" IEEE Vehicular Technology Conference, spring 2003, page(s): 1979 - 1983 vol.3.

[8] Harri Holma and Antti Toskala, "WCDMA for UMTS-Radio Access Third Generation Mobile Communications"- John Wiley & Sons.

[9] Jaana Laiho, Achim Wacker, Tomas Novosad, "Radio Network Planning and Optimisation for UMTS"-Second Edition John Wiley & Sons.

[10] A. Skopljak, "Multiantenna CDMA systems and their usage in 3G network", University of Sarajevo, 2007.

[11] Faruque, Saleh, "Cellular Mobile System Engineering", Artech House Publishers, 1996.

[12] Jhong Sam Lee, Leonard E. Miller, "CDMA Systems Engineering Handbook"-Artech House Publishers.

Appendix

Algorithm for Capacity Analysis Using Sectorization:

Begin
Set energy per bit to noise spectral density ratio (Eb/No) = [1 2 3 4 5 6 7 8 9 10 16 20]
Set soft handover gain (H) factor = [0.1 1 1.5 2 3]
Set inter-cell interference (β) = [.1 1.2 1.55 2]
Set channel activity for data (α) = [1]
Set channel activity for voice (α) = [0.1 .3 .38 0.7 0.9]
Set thermal noise (η) in (20 Kelvin) dbm/Hz= [-173.93]
Set user signal power (S1) in dbm= [21]
Set shadow fading (sh_fd) in db = [8]
Set cell range in Km (R_{cell}) = [2]
Set chip rate (W) = [3840000]
Set base band information rate in Kbps (R) = [12.2 64 144 384 2000]
Set base antenna height in meter (h_b) = [20]
Set user antenna height in meter (h_{UE}) = [2]
Set sector (D) = [1 2 3 4 6]
Set frequency range in MHz (f_c) = [2000]
Set data rate in Kbps (R) = [12.2 64 144 384 2000]
Set array antenna gain (A_g) in db= [1 2 3.5 5]
//Processing
Processing gain (PG) = 10log (W/R)
Propagation loss in dense urban (Pro_loss) =

$$46.3+33.9\log(f_c)-13.82\log h_b-3.2[\log(11.75_{\mathrm{UE}})]^2+4.97+(44.9-6.55\log h_b)\log d+3$$

Signal Power (S) = S1-Pro_loss-sh_hd
//Output
Number of Users $N_s=1+(\frac{W/R}{E_b/N_o}-\frac{\eta}{S})\frac{1}{(1+\beta)\alpha}\times D\times H\times A_g$

End

Algorithm for Coverage and Data rates Analysis Using Sectorization:

Begin
Set Transmitter=User Equipment
Set Receiver=Node B
Set mobile max power in dbm (mo_mx) = [21]
Set mobile gain in db (M_G) = [0]
Set cable and connector losses in db (ca_cn_loss) = [3]
Set thermal noise in dbm/Hz (η) = [-173.93]
Set node B noise figure in db (nodeB_NF) = [5]
Set target load (tar_ld) = [.4]
Set chip rate (W) = [3840000]
Set base antenna height in meter (h_b) = [20]
Set user antenna height in meter (h_{UE}) = [2]
Set energy per bit to noise spectral density ratio (Eb/No) = [5]
Set Power Control Margin or Fading Margin (MPC) = [4]
Set Value for sectors (Sec) = [1 2 3 4]
Set constant value for sectors (K) = [2.6 1.6 1.95 2.6]
Set data rate in Kbps (R) = [100 200 300 400 500 600 2000]
//Processing
Chip rate in (W) = [3840000]
Processing gain (PG) = (W/R)
Effective isotropic radiated power (EIRP) = mo_mx-ca_cn_loss+M_G
Node B noise density (nodeB_ND) = η+ nodeB_NF
Node B noise power (nodeB_NPW) = nodeB_ND +W_db
Interference margin (IM) =-10log (1-tar_ld)

Node B Interference Power (nodeB_IP) = $10\log\left(10^{(noisepower+InterfrenceM\arg in)/10} - 10^{Noisepower/10}\right)$

Node B Noise and interference (nodeB_NIFPW) = $10\log\left(10^{(noisepower)/10} - 10^{(Interfrencepower)/10}\right)$

Node B antenna gain (NodeB_AG) in db = [18]

Receiver Sensitivity (S_{rx}) = Eb/No-PG+ nodeB_NIFPW

Total Allowable Path loss=EIRP-S_{rx}+ nodeB_AG-MPC=EIRP-(E_b/N_o-PG+ nodeB_NIFPW) + NodeB_AG-MPC

Path loss in dense urban (Durban_Ploss) =

$$46.3 + 33.9\log(f_c) - 13.82\log h_b - 3.2[\log(11.75_{\mathrm{UE}})]^2 + 4.97 + (44.9 - 6.55\log h_b)\log d + 3$$
$$= 142.17 + 36.37\log d$$

//Output

Cell radius (d) = 10^ ((1/36.37)* (EIRP-(Eb/No-PG+ nodeB_NIFPW) + nodeB_AG-MPC-142.17))

Cell Area (A) = K*d2

End

Architecture Aware Programming on Multi-Core Systems

M.R. Pimple
Department of Computer Science & Engg.
Visvesvaraya National Institute of Technology,
Nagpur – 440010 (India)

S.R. Sathe
Department of Computer Science & Engg.
Visvesvaraya National Institute of Technology,
Nagpur – 440010 (India)

***Abstract*— In order to improve the processor performance, the response of the industry has been to increase the number of cores on the die. One salient feature of multi-core architectures is that they have a varying degree of sharing of caches at different levels. With the advent of multi-core architectures, we are facing the problem that is new to parallel computing, namely, the management of hierarchical caches. Data locality features need to be considered in order to reduce the variance in the performance for different data sizes. In this paper, we propose a programming approach for the algorithms running on shared memory multi-core systems by using blocking, which is a well-known optimization technique coupled with parallel programming paradigm, OpenMP. We have chosen the sizes of various problems based on the architectural parameters of the system like cache level, cache size, cache line size. We studied the cache optimization scheme on commonly used linear algebra applications – matrix multiplication (MM), Gauss-Elimination (GE) and LU Decomposition (LUD) algorithm.**

Keywords- multi-core architecture; parallel programming; cache miss; blocking; OpenMP; linear algebra.

I. INTRODUCTION

While microprocessor technology has delivered significant improvements in clock speed over the past decade, it has also exposed a variety of other performance bottlenecks. To alleviate these bottlenecks, microprocessor designers have explored alternate routes to cost effective performance gains. This has led to use of multiple cores on a die. The design of contemporary multi-core architecture has progressively diversified from more conventional architectures. An important feature of these new architectures is the integration of large number of simple cores with software managed cache hierarchy with local storage. Offering these new architectures as general-purpose computation platforms creates number of problems, the most obvious one being programmability. Cache based architectures have been studied thoroughly for years leading to development of well-known programming methodologies for these systems, allowing a programmer to easily optimize code for them. However, multi-core architectures are relatively new and such general directions for application development do not exist yet.

Multi-core processors have several levels of memory hierarchy. An important factor for software developers is how to achieve the best performance when the data is spread across local and global storage. Emergence of cache based multi-core systems has created a "cache aware" programming consensus. Algorithms and applications implicitly assume the existence of a cache. The typical example is linear algebra algorithms. To achieve good performance, it is essential that algorithms be designed to maximize data locality so as to best exploit the hierarchical cache structures. The algorithms must be transformed to exploit the fact that a cache miss will move a whole cache-line from main memory. It is also necessary to design algorithms that minimize I/O traffic to slower memories and maximize data locality. As the memory hierarchy gets deeper, it is critical to efficiently manage the data. A significant challenge in programming these architectures is to exploit the parallelism available in the architecture and manage the fast memories to maximize the performance. In order to avoid the high cost of accessing off-chip memory, algorithms and scheduling policies must be designed to make good use of the shared cache[12]. To improve data access performance, one of the well-known optimization technique is tiling[3][10]. If this technique is used along with parallel programming paradigm like OpenMP, considerable performance improvement is achieved. However, there is no direct support for cache aware programming using OpenMP for shared memory environment. Hence, it is suggested to couple OpenMP with tiling technique for required performance gain.

The rest of the paper is organized as follows. Section II describes the computing problem which we have considered. The work done in the related area is described in section III. Implementation of the problems is discussed in section IV. Experimental setup and results are shown in section V. The performance analysis is carried out in section VI.

II. COMPUTING PROBLEM

As multi-core systems are becoming popular and easily available choice, for not only high performance computing world but also as desktop machines, the developers are forced to tailor the algorithms to take the advantage of this new platform. As the gap between CPU and memory performance continues to grow, so does the importance of effective utilization of the memory hierarchy. This is especially evident in compute intensive algorithms that use very large data sets, such as most linear algebra problems. In the context of high performance computing world, linear algebra algorithms have to be reformulated or new algorithms have to be developed in order to take advantage of the new architectural features of these new processors. Matrix factorization plays an important role in a large number of applications. In its most general form, matrix factorization involves expressing a given matrix as a product of two or more matrices with certain properties. A large number of matrix factorization techniques have been proposed and researched in the matrix computation literature to meet the requirements and needs arising from different application domains. Some of the factorization techniques are categorized into separate classes depending on whether the original matrix is dense or sparse. The most commonly used matrix factorization techniques are LU, Cholesky, QR and singular value decomposition (SVD).

The problem of dense matrix multiplication (MM) is a classical benchmark for demonstrating the effectiveness of techniques that aim at improving memory utilization. One approach towards the cache effective algorithm is to restructure the matrices into sequence of tiles. The copying operation is then carried out during multiplication. Also, for a system AX=B, there are several different methods to obtain a solution. If a unique solution is known to exist, and the coefficient matrix is full, a direct method such as Gaussian Elimination(GE) is usually selected.

LU decomposition (LUD) algorithm is used as the primary means to characterize the performance of high-end parallel systems and determine its rank in the Top 500 list[11]. LU Factorization or LU decomposition is perhaps the most primitive and the most popular matrix factorization techniques finding applications in direct solvers of linear systems such as Gaussian Elimination. LU factorization involves expressing a given matrix as product of a lower triangular matrix and an upper triangular matrix. Once the factorization is accomplished, simple forward and backward substitution methods can be applied to solve a linear system. LU factorization also turns out to be extremely useful when computing the inverse or determinant of a matrix because computing the inverse or the determinant of a lower or an upper triangular matrix is relatively easy.

III. Related Work

Since multi-core architectures are now becoming mainstream, to effectively tap the potential of these multiple units is the major challenge. Performance and power characteristics of scientific algorithms on multi-core architectures have been thoroughly tested by many researchers[7]. Basic linear algebra operations on matrices and vectors serve as building blocks in many algorithms and software packages. Loop tiling is an effective optimization technique to boost the memory performance of a program. The tile size selection using cache organization and data layout, mainly for single core systems is discussed by Stephanie Coleman and Kathryn S. Mckinley [10].

LU decomposition algorithm decomposes the matrix that describes a linear system into a product of a lower and an upper triangular matrix. Due to its importance into scientific computing, it is well studied algorithm and many variations to it have been proposed, both for uni and multi-processor systems. LU algorithm is implemented using recursive methods [5], pipelining and hyperplane solutions [6]. It is also implemented using blocking algorithms on Cyclops 64 architecture [8]. Dimitrios S. Nikolopoulos, in his paper [4] implemented dynamic blocking algorithm. Multi-core architectures with alternative memory subsystems are evolving and it is becoming essential to find out programming and compiling methods that are effective on these platforms. The issues like diversity of these platforms, local and shared storage, movement of data between local and global storage, how to effectively program these architectures; are discussed in length by Ioannis E. Venetis and Guang R. Gao [8]. The algorithm is implemented using block recursive matrix scheme by Alexander Heinecke and Michael Bader [1]. Jay Hoeflinger, Prasad Allavilli, Thomas Jackson and Bob Kuhn have studied scalability issues using OpenMP for CFD applications[9]. OpenMP issues in the development of parallel BLAS and LAPACK libraries have also been studied[2]. However, the issues, challenges related with programming and effective exploitation of shared memory multi-core systems with respect to cache parameters have not been considered.

Multi-core systems have hierarchical cache structure. Depending upon the architecture, there can be two or three layers, with private and shared caches. When implementing the algorithm, on shared memory systems, cache parameters must be considered. The tile size selection for any particular thread running on a core is function of size of L_1 cache, which is private to that core as well as of L_2 cache which is a shared cache. If cache parameters like, cache level, cache size, cache line size are considered, then substantial performance improvement can be obtained. In this paper, we present the parallelization of MM, GE and LUD algorithm on shared memory systems using OpenMP.

IV. Implementation

In this paper we have implemented parallelization of most widely used linear algebra algorithms, matrix multiplication, gauss elimination and LU decomposition, on multi-core systems. Parallelization of algorithms can also be implemented using message passing interface (MPI). Pure MPI model assumes that, message passing is the correct paradigm to use for all levels of parallelism available in the application and that the application “topology” can be mapped efficiently to the hardware topology. However, this may not be true in all cases. For matrix multiplication problem, the data can be decomposed into domains and these domains can be independently passed to and processed by various cores. While, in case of LU decomposition or GE problem, task dependency prevents to distribute the work load independently to all other processors. Since the distributed processors do not share a common memory subsystem, the computing to communication ratio for this problem is very low. Communication between the processors on the same node goes through the MPI software layers, which adds to overhead. Hence, pure MPI implementation approach is useful when domain decomposition can be used; such that, the total data space can be separated into fixed regions of data or domains, to be worked on separately by each processor.

For GE and LUD problems, we used the approach of 1D partitioning of the matrix among the cores and then used OpenMP paradigm for distributing the work among number of threads to be executed on various cores. The approach of 2 D partitioning of data among cores is more suitable for array processors. For a shared memory platform, all the cores on a single die share the same memory subsystem, and there is no direct support for binding the threads to the core using OpeMP. So, we restricted our experiments with 1D partitioning technique and applied parallelization for achieving speedup using OpenMP.

A. Architecture Aware Parallelization

To cope up with memory latency, all data required during any phase of the algorithm are made available in the cache. The data sets so chosen, are accommodated into the cache. Considering the cache hierarchy, the tile size selection depends upon cache size, cache line size to eliminate self interference misses. Now depending upon the architecture of the underlying machine, the computation work is split into number of cores available. One dimensional partitioning of data is done, so that, every core receives specific number of rows (or columns), such that, the data fits in the shared cache. The blocking technique is then used which ensures that the maximum block size is equal to the size of private cache belonging to the core. Parallel computation is carried out using OpenMP pragmas by individual cores.

B. Determining Block Size

In order to exploit cache affinity, the block size is chosen such that, the data can be accommodated into the cache. The experiments were carried out on square matrix of size N. Let 's' be the size of each element of matrix and 'C_s' be the size of shared cache. Let the block size be $B \times B$.

1. For blocked matrix multiplication, C = A x B, block of matrix A & B, and one row of matrix C should be accommodated into the cache. Then the required block size can be calculated using :

$$2sB^{2} + sB = C_s$$

For large cache size, we get,

$$B = \sqrt{C_s/2s} \quad (1)$$

2. For GE problem, the size of input matrix is $[N]x[N+1]$. The required block size can be calculated with the following equation:

$$B \times (B+1) \times s = C_s$$

So, the optimal block size,

$$B = \sqrt{C_s/s} \quad (2)$$

3. For LU decomposition algorithm with same matrix used for storing lower and upper triangular matrix, the optimal block size comes out to be

$$B = \sqrt{C_s} \quad (3)$$

C. Effect of Cache Line Size

Let cache line size be C_{ls}. Without the loss of generality, we assume that the first element of input array falls in the first position of cache. The number of rows that completely fit in the cache can be calculated as :

$$\text{Rows} = C_s/C_{ls} \quad (4)$$

For every memory access, the entire cache line is fetched. So block size $B < C_{ls}$ will lead to self interference misses. Also if $B > C_{ls}$, system will fetch additional cache lines, which may in turn lead to capacity misses; as less number of rows can be accommodated in the cache. So to take the advantage of spatial locality, the block sizes chosen were integral multiple of cache line size C_{ls}. We assumed that, every row in the selected tile is aligned on a cache line boundary. After finding the row size, block size can be calculated.

Block Size B = k x C_{ls}

if $(B \bmod C_{ls} = 0)$ (*k is integer*)

Or $B = \left\lfloor \frac{Rows}{C_{ls}} \right\rfloor \times C_{ls}$

Maximum Speed up is achieved when -

$(B \propto C_{ls})$ or

when B is multiple of number of Rows

The algorithm for block size selection is presented in Fig. 1.

Further improvement in the performance is achieved by using the technique of register caching for the array elements, that are outside the purview of the "*for*" loop (like value a[i][j] shown in Fig. 3). This value is cached, which is then shared by all the threads executing the "*for*" loop. The OpenMP implementation of matrix multiplication and GE problem is given in Fig. 2 and Fig. 3 respectively.

D. LU Decomposition

The main concept is to partition the matrix into smaller blocks with a fixed size. The diagonal entry in each block is processed by master thread on a single core. Then for calculating the entries in the upper triangular matrix, each row is partitioned into number of groups equal to number of cores; so that each group is processed by each core. Similarly, for calculating the entries in the lower triangular matrix, each column is partitioned into number of groups equal to number of cores; so that each group is processed by each core. The implementation divides the matrix into fixed sized blocks, that fit into the L1 data cache of the core creating first level of memory locality. On the shared memory architecture, the whole matrix is assumed to be in the globally accessible memory address space. The algorithm starts by processing the diagonal block on one processor, while all other processors wait for the barrier. When this block finishes, the blocks on the same row are processed by various cores in parallel. Then the blocks on same column are processed by various cores in parallel. In turn, each processor waits for the barrier again for the next diagonal block.

The storage space can further be reduced by storing lower and upper triangular matrices in a single matrix. The diagonal elements of lower triangular matrix are made 1, hence, they need not be stored. But this method suffers from the problem of load imbalance, if number of elements processed in each row or column by each core is not divisible by number of cores available. Also, the active portion of the matrix is reduced after each iteration and hence, load allocation after each iteration is not constant.

1) LUD computation:

Let A be an $n \times n$ matrix with rows and columns numbered from 0 to (n-1). The factorization consists of n major steps. Each step consisting of an iteration of the outer loop starting at line 3 of Fig. 5. In step k, first the partial column $A[k+1:n,\ k]$ us divided by $A[k,k]$. Then the outer product $A[k+1:n,\ k]$ x $A[k,\ k+1:n]$ is subtracted from the $(n-k)$ x $(n-k)$ sub matrix $A[k+1:n,\ k+1:n]$. For each iteration of the outer loop $k = 0\ to\ (n-1)$, the next nested loop in the algorithm goes from $k+1\ to\ (n-1)$.

```
Procedure BS(CS, CLS,N,B)
Input: CS: Cache Size
CLS: Cache Line Size
s: Size of each element in input
N: Input Matrix Rows
Output : B : Block Size(square)
Total cache lines = CS / CLS
No of rows (NR) from input problem size that can be
accommodated in cache
NR = ( )
The optimal block size B
If (NR > CLS)
       B= k xCLS   // Where k is integer constant
Else (B = CLS)
```

Figure 1. Block Size Selection

```
void mat-mult()  // matrix multiplication //
{ for (i=0; i<N; i=i+B)
 {  Read block of a & c;
    Read block ofbB;
    omp_set_num_threads(Omp_get_num_proc());
    #pragma omp parallel for shared(a,b,c,i)
          private(r,i1,j) schedule (static)
    for (r=i; r<(min(i+B, N)); r++)
     for (i1=i; i1<(min(i+B,N)); i1++)
       {  for(j=0; j<N; j++)
                   c[r][i1] += a[r][j] * b[j][i1];
       }
    Write block of c ;
 }}
```

Figure 2. Parallel Matrix Multiplication

A typical computation of LU factorization procedure in the k^{th} iteration of the outer loop is shown in Fig. 4. The k^{th} iteration of the outer loop does not involve any computation on rows 1 to (k-1) or columns 1 to (k-1). Thus at this stage, only the lower right (n-k) x (n-k) sub matrix of A is computationally active. So the active part of the matrix shrinks towards the bottom right corner of the matrix as the computation proceeds.

The amount of computation increases from top left to bottom right of the matrix. Thus the amount of work done differs for different elements of matrix. The work done by the processes assigned to the beginning rows and columns would be far less than those assigned to the later rows and columns. Hence, static scheme of block partitioning can potentially lead to load imbalance. Secondly, the process working on a block may idle even when there are unfinished tasks associated with that block.

```
void forwardSubstitution()  // GaussElimination loop//
// Matrix size (n x n)
{  int i, j, k, max, kk, p, q; float t;
    for (i = 0; i < n; ++i)
   {  max = i;
     for (j = i + 1; j < n; ++j)
    if (a[j][i] > a[max][i])  max = j;
     for (j = 0; j < n + 1; ++j)
    {   t = a[max][j]; a[max][j] = a[i][j];
         a[i][j] = t;
    }
   for (j = n; j >= i; --j)
   {   for (k=i+1; k<n; k=k+B)
      { x=a[i][i];   // Register Caching //
       #pragma omp parallel for shared(i,j,k) private(kk)
schedule (static)
      for (kk=k; kk<(min(k+B, n)); ++kk)
             a[kk][j] -= a[kk][i]/x * a[i][j];
      }
   }
}
```

Figure 3. OpenMP parallelization of GE loop

This idling can occur if the constraints imposed by the task-dependency graph do not allow the remaining tasks on this process to proceed until one or more tasks mapped onto other processes are completed.

2) LUD OpenMP parallelization:

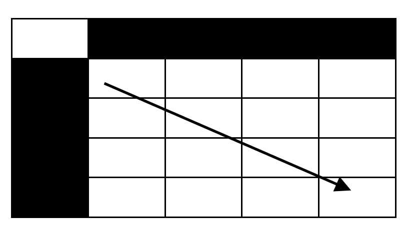
For parallelization of LU decomposition problem on shared memory, we used tiling technique with OpenMP paradigm. The block size B is selected such that, the matrix size is accommodated in a shared cache. The actual data block used by each core is less than the size of private cache so that locality of memory access for each thread is maintained.

For LUD algorithm, due to the task dependency at each iteration level, the computation cannot be started simultaneously on every core. So, algorithm starts on one core. Diagonal element is executed by master core. After the synchronization barrier, the computation part of non-diagonal elements is split over the available cores.

After computing a row and column of result matrix, again the barrier is applied to synchronize the operations for the next loop. The size of data computed by each core is determined by block size.

The size of data dealt by each core after each iteration is not the same. With static scheduling, the chunk is divided exactly into the available multiple threads and every thread works on the same amount of data.

Fig. 5 illustrates the OpenMP parallelization. The size of input matrix ‘a’ is ‘N’.



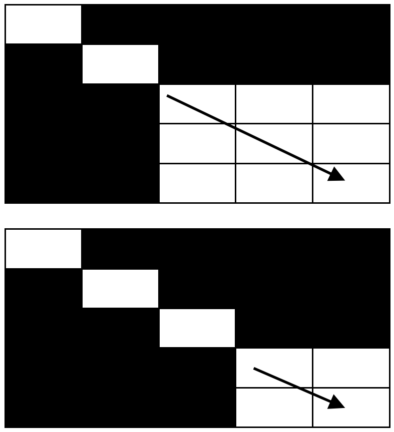


Figure 4.Processing of blocks of LU Decomposition

```
Lu-Fact (a)
{
  for (k=0; k<N; k++)
  {   #pragma omp single
      for(j=k+1; j<(N); j++)
          a[j][k]=a[j][k]/a[k][k];
     #pragma omp parallel for shared (a,k)
 private(i) schedule static
      for(j=k+1; j<(N); j=j+B)
      for (jj=j; jj<min(jj+B, N), jj++)
     {    v=a[k][jj];        --- caching the value
          #pragma omp parallel for shared
 (a,k,jj) private(i) schedule static
          for(i=k+1; i<(N); i++)
              a[i][jj]= a[i][jj]- (a[i][k]*v);
     }
   }
 }
```

Figure 5. OpenMP implementation of LU Decomposition algorithm

V. Experimental Setup & Results

We conducted the experiments to test cache aware parallelization of MM, GE, LUD algorithms on Intel Dual core, 12 core and 16 core machines. Each processor had hyper threading technology such that, each processor can execute simultaneously instructions from two threads, which appear to the operating system as two different processors and can run a multi program workload. The configuration of the systems is given in Table 1.

Each processor had 32 KB data cache as L_1 cache. Intel Xeon processors (12 & 16 cores) had an eight way set associative 256 KB L_2 cache and 12 MB L_3 cache dynamically shared between the threads. The systems run Linux 2.6.x Blocked LU decomposition was parallelized at two levels using OpenMP.

We used relatively large data sets, so that the performance of the codes becomes more bound to the L_2 and L_3 cache miss latencies. The programs were compiled with C compiler (gcc 4.3.2). Fig. 6 and Fig. 7 show the speed up achieved when the block sizes are such that, the data fits in L_2 cache for matrix multiplication and Gauss elimination algorithm respectively. Fig. 8 and Fig. 9 show the results of LU decomposition algorithm for various matrix sizes on dual and 16 core system respectively

Table 1. System configuration

Processors	**Intel(R) Core™2 Duo CPU E7500**	**Intel(R) Dual Core CPU E5300**	**Intel(R) Xeon(R) CPU X5650 (12 cores)**	**Intel(R) Xeon(R) CPU E5630 (16 cores)**
Core frequency	2.93 GHz	2.60 GHz	2.67 GHz	2.53 GHz
L1 Cache size	32 KB I cache, 32 KB D cache	32 KB I cache, 32 KB D cache	32 KB I cache, 32 KB D cache	32 KB I cache, 32 KB D cache
L2 Cache size	3072 KB, shared	2048 KB, Shared	256 KB	256 KB
L3 Cache size	---	---	12 MB	12 MB

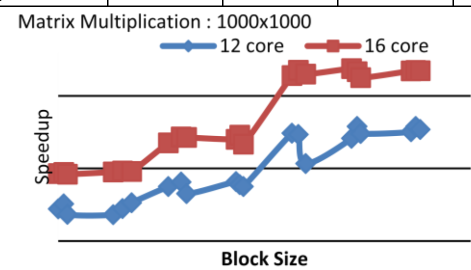


Figure 6. Matrix Multiplication: Data fits in L_2 Cache

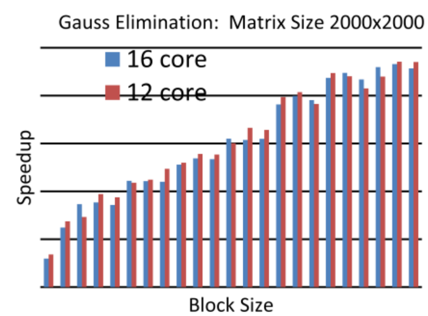


Figure 7. Gauss Elimination on 12 & 16 core machines, Data accommodated in L_2 cache

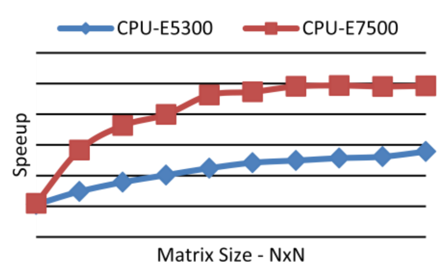


Figure 8. Speedup for LU Decomposition on Dual core CPU

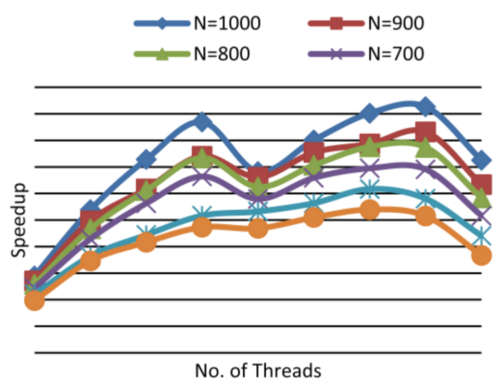


Figure 9. Speedup for LU Decomposition on 16 core machine
(Matrix size- NxN)

VI. Performance Analysis

The strategy of parallelization is based on two observations. One is that the ratio computation to communication should be very high for implementation on shared memory multi-core systems. And second is that the memory hierarchy is an important parameter to be taken into account for the algorithm design which affects load and store times of the data. Considering this, we implemented the algorithms matrix multiplication and Gauss elimination with a blocking scheme that divides the matrices into relatively small square tiles. The optimal block size is selected for each core, such that the tile is accommodated in the private cache of each core and thus avoids the conflict misses. This approach of distributing the data chunks to each core greatly improves the performance. Fig. 6 and Fig. 7 shows the performance improvement when block size is multiple of cache line size. Whenever block size is greater or less than the cache line size, performance suffers. This is due to reloading overheads of entire new cache line for the next data chunk. With this strategy, we got the speedup of 2.1 on 12 core machine and speed up of 2.4 on 16 core machine. The sub linear speedups in Fig. 6 and 7 for lower block sizes are attributed to blocking overheads.

For Gauss elimination and LU decomposition problem, the OpenMP pragma, splits the data among the available cores. The size of data dealt by every core, after every iteration is different. This leads to load imbalance problem. The chunk scheduling scheme, demands the chunk calculations at every iteration and hence affects performance. However, static scheduling ensures equal load to every thread and hence reduces the load imbalance. For LU decomposition problem with 1D partitioning of data among the cores, we observed a speedup of 1.39, & 2.46 for two dual core machines and speedup of 3.63 on 16 core machine. The maximum speedup is observed when the number of threads is equal to the number of (hardware) threads supported by the architecture. Fig. 9 shows the speed up when 16 threads are running on a 16 core machine. Speed up is directly proportional to the number of threads. The performance degrades when more software threads are in execution than the threads supported by architecture. So, for 18 threads, scheduling overhead increases and performance is degraded. However, when number of threads is more than 8, performance degrades due to communication overheads. This is because, 16 core Intel Xeon machine comprises of 2 quad cores connected via QPI link. Fig. 9 shows performance enhancement up to eight threads and degradation in the performance when number of threads is ten. When the computation is split across all the available sixteen threads, speed up is again observed, where communication overhead is amortized over all cores. Further enhancement in the performance is achieved when method of register caching is used for loop independent variables in the program. Many tiling implementations do not consider this optimal block size considerations with cache attributes. However, our implementation considers the hierarchy of caches, cache parameters and arrives at optimal block size. The block size calculations are governed by the architecture of the individual machine and the algorithm under consideration. Once the machine parameters and input problem size is available, the tailoring of the algorithm accordingly improves the performance to a greater extent. Of course, there is a significant amount of overhead in the OpenMP barriers at the end of loops; which means that load imbalance and not data locality is the problem.

VII. Conclusion & Future Work

We evaluated performance effects of exploiting architectural parameters of the underlying platform for programming on shared memory multi-core systems. We studied the effect of private cache L_1, shared cache L_2, cache line size on execution of compute intensive algorithms. The effect of exploiting L1 cache affinity does not affect the performance much, but the effects of exploiting L2 cache affinity is considerable, due its sharing among multiple threads and high reloading cost for larger volumes. If these factors are considered and coupled with parallel programming paradigm like OpenMP, performance enhancement is achieved. We conclude that, affinity awareness in compute intensive algorithms on multi-core systems is absolutely essential and will improve the performance significantly. We plan to extend the optimization techniques for the performance enhancement on multi-core systems by considering the blocking technique at register level and instruction level. We also plan to investigate and present generic guide lines for compute intensive algorithms on various multi-core architectures.

References

[1] Alexander Heinecke and Michael Bader, "Towards many-core implementation of LU decomposition using Peano Curves," *UCHPC-MAW'09*, May 18-20, 2009, Ischia, Itali.

[2] C. Addison, Y. Ren, and M. van Waveren, "OpenMP issues arising in the development of Parallel BLAS and LAPACK libraries," In *Scientific Programming Journal*, Volume 11, November 2003, pages: 95-104,IOS Press.

[3] Chung-Hsing Hsu, and Ulrich Kremer, "A quantitative analysis of tile size selection algorithms," *The Journal of Supercomputing*, 27, 279-294, 2004.

[4] Dimitrios S. Nikolopoulos, "Dynamic tiling for effective use of shared caches on multi-threaded processors," *International Journal of High*

Performance Computing and Networking, 2004, Vol. 2, No. 1, pp. 22-35.

[5] F.G. Gustavson, "Recursion leads to automatic variable blocking for dense linear algebra algorithms," *IBM Journal of Research and Development*, 41(6):737-753, Nov.,1997.

[6] H. Jin, M. Frumkin, and J. Yan, "The OpenMP implementation of NAS parallel benchmarks and its performance," Technical report nas-99-011, NASA Ames Research Centre, 1999.

[7] Ingyu Lee, "Analyzing performance and power of multi-core architecture using multithreaded iterative solver," *Journal of Computer Science*, 6 (4): 406-412, 2010.

[8] Ioannis E. Venetis, and Guang R. Gao, "Mapping the LU decomposition on a many-core architecture: Chanllenges and solutions," *CF'09*, May 18-20, 2009, Ischia, Itali.

[9] Jay Hoeflinger, Prasad Alavilli, Thomas Jackson, and Bob Kuhn, "Producing scalable performance with OpenMP: Experiments with two CFD applications," *International Journal of Parallel computing*. 27(2001), 391-413.

[10] Stephanie Coleman, and Kathryn S. McKinley, "Tile size selection using cache organization and data layout," *Proceedings of the ACM SIGPLAN Conference on Programming Language Design & Implementation*, California, United States, 1995, pages: 279-290

[11] The Top 500 List. http://www.top500.org

[12] Vahid Kazempour, Alexandra Fedorova, and Pouya Alagheband, "Performance implications of cache affinity on multicore processors," *Proceedings of the 14th International Euro-Par Conference on Parallel Processing*, Spain, 2008, pages:151-161

AUTHORS PROFILE

S.R. Sathe received M.Tech. in Computer Science from IIT, Bombay (India) and received Ph.D. from Nagpur University(India). He is currently working as a Professor in the Department of Computer Science & Engineering at Visvesvaraya National Institute of Technology, Nagpur(India). His research interests include paralle and distributed systems, mobile computing and algorithms.

M.R. Pimple is working as a Programmer in the Department of Computer Science & Engineering at Visvesvaraya National Institute of Technology, Nagpur(India) and persuing her M.Tech. in Computer Science. Her area of interests are computer architecture, parallel processing.

Implementation of ISS - IHAS (Information Security System – Information Hiding in Audio Signal) model with reference to proposed e-cipher Method

Prof. R. Venkateswaran
Department of CA, Nehru College of Management,
Research Scholar-Ph.D, Karpagam University
Coimbatore, TN, INDIA

Dr. V. Sundaram, Director
Department of CA, Karpagam College of Engineering
Affiliated to Anna University of Technology,
Coimbatore, TN, INDIA

***Abstract*— This paper shows the possibility of exploiting the features of E- cipher method by using both cryptography and Information hiding in Audio signal methods used to send and receive the message in more secured way. Proposed methodology shows that successfully using these Poly substitutions methods (Proposed E-Cipher) for Encode and decodes messages to evolve a new method for Encrypting and decrypting the messages.**

Embedding secret messages using audio signal in digital format is now the area in focus. There exist numerous steganography techniques for hiding information in audio medium.

In our proposed theme, a new model ISS-IHAS - Embedding Text in Audio Signal that embeds the text like the existing system but with strong encryption that gains the full advantages of cryptography. Using steganography it is possible to conceal the full existence of the original text and the results obtained from the proposed model is compared with other existing techniques and proved to be efficient for textual messages of minimum size as the size of the embedded text is essentially same as that of encrypted text size. This emphasis the fact that we are able to ensure secrecy without an additional cost of extra space consumed for the text to be communicated.

Keywords- Encryption; Decryption; Audio data hiding; Mono Substitution; Poly Substitution.

I. Objectives of the Project

The main purpose of Audio steganography is to hide a message in some cover media, to obtain new data, practically indistinguishable from the original message, by people, in such a way that an eavesdropper cannot detect the presence of original message in new data. With computers and Networks, there are many other ways of hiding information, such as Covert channels, Hidden text within WebPages Hiding files in “Plain sight”, Null ciphers.

Today, the internet is filled with tons of programs that use steganography to hide the secret information. There are so many medias are used for digitally embedding message such as plaintext, hypertext, audio/video, still image and network traffic. There exists a large variety of steganographic techniques with varying complexity and possessing some strong and weak aspects.

Hiding information in text is the most popular method of Steganography. It is used to hide a secret message in every nth character or altering the amount of white space after lines or between words of a text message [1]. It is used in initial decade of internet era. But it is not used frequently because the text files have a small amount of redundant data. But this technique lacks in payload capacity and robustness. To hide data in audio files, the secret message is embedded into digitized audio signal. Audio data hiding method provides the most effective way to protect privacy. Key aspect of embedding text in audio files is that, no extra bytes are generated for embedding. Hence it is more comfortable to transmit huge amount of data using audio signal. Embedding the secret messages in digital sound is usually a very difficult process [2].

II. Proposed ISS – IHAS Model

The following IHAS Model provides a very basic description of the audio steganographic process in the sender side and receiver side.

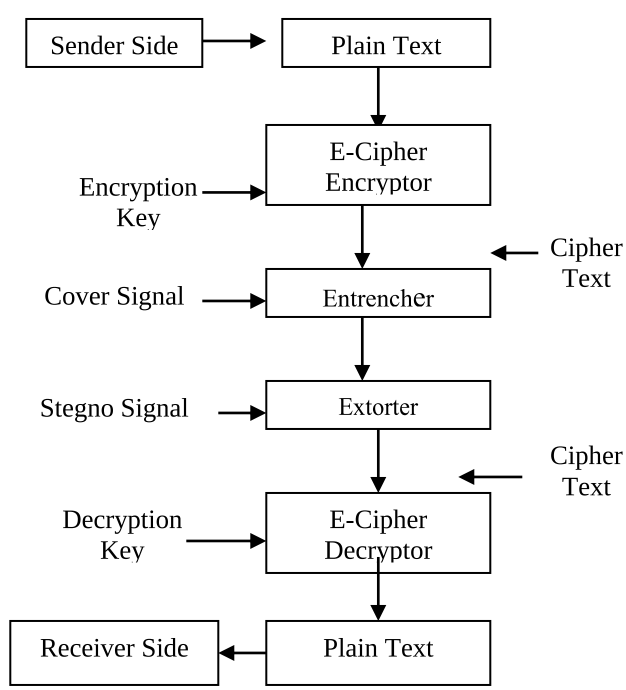


Fig 2.1 System Flow – ISS- IHAS model

The original text encrypted by E-Cipher using an encryption key. The model implements E-Cipher encryption as it proves to be more efficient. The encrypted text is passed on to Entrencher that embeds the encrypted text inside the cover signal which is in audio format *.wav resulting in stego signal. This process happens at sender side. This stego signal is communicated using Network medium. At the receiver side the stego signal is passed on to Extorter module that extracts embedded text from the audio signal that was used a s cover medium,. The resultant cipher text is then decrypted using E-cipher Decryptor module. The final plain text can then be used for further processing.

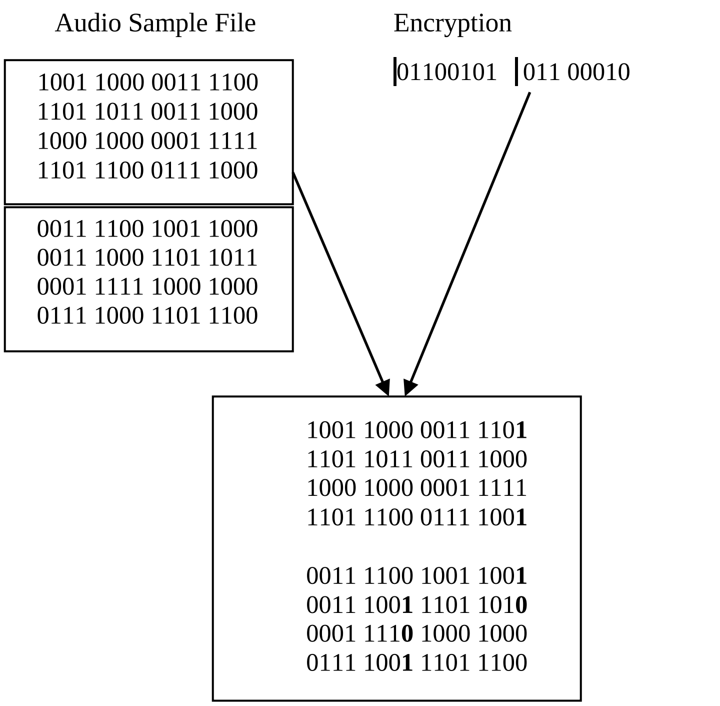


Fig. 2.2 ISS- IHAS encoding format

To hide a letter A & B to an digitized audio file where each sample is represented with 16 bits then the LSB bit of sample audio file is replaced with each bit of binary equivalent of the letter A & B[4].

III. PERSPECTIVE STUDY ON VARIOUS METHODS

In audio steganography, secret message is embedded into digitized audio signal which result slight altering of binary sequence of the corresponding audio file. There are several methods are available for audio steganography [3]. Some of them are as follows: -

LSB Coding:

Least significant bit (LSB) coding is the simplest way to embed information in a digital audio file. By substituting the least significant bit of each sampling point with a binary message, LSB coding allows for a large amount of data to be encoded. The following diagram illustrates how the message 'HEY' is encoded in a 16-bit CD quality sample using the LSB method:

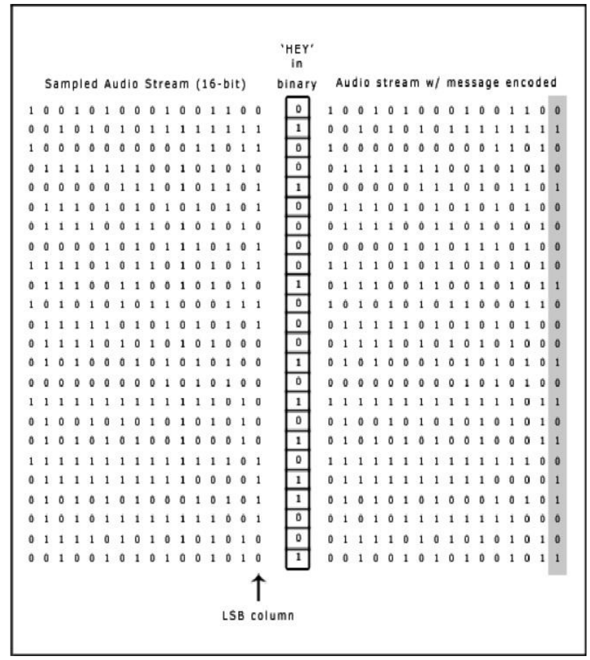


Fig.3.1. Message 'HEY' is encoded in a 16-bit CD quality sample using the LSB method

In LSB coding, the ideal data transmission rate is 1 kbps per 1 kHz. In some implementations of LSB coding, however, the two least significant bits of a sample are replaced with two message bits. This increases the amount of data that can be encoded but also increases the amount of resulting noise in the audio file as well. Thus, one should consider the signal content before deciding on the LSB operation to use. For example, a sound file that was recorded in a bustling subway station would mask low-bit encoding noise. On the other hand, the same noise would be audible in a sound file containing a piano solo [9].

To extract a secret message from an LSB encoded sound file, the receiver needs access to the sequence of sample indices used in the embedding process. Normally, the length of the secret message to be encoded is smaller than the total number of samples in a sound file. One must decide then on how to choose the subset of samples that will contain the secret message and communicate that decision to the receiver. One trivial technique is to start at the beginning of the sound file and perform LSB coding until the message has been completely embedded, leaving the remaining samples unchanged. This creates a security problem, however in that the first part of the sound file will have different statistical properties than the second part of the sound file that was not modified. One solution to this problem is to pad the secret message with random bits so that the length of the message is equal to the total number of samples. Yet now the embedding process ends up changing far more samples than the transmission of the secret required. This increases the probability that a would-be attacker will suspect secret communication [8].

Parity Coding

Instead of breaking a signal down into individual samples, the parity coding method breaks a signal down into separate regions of samples and encodes each bit from the secret message in a sample region's parity bit. If the parity bit of a selected region does not match the secret bit to be encoded, the process flips the LSB of one of the samples in the region. Thus, the sender has more of a choice in encoding the secret bit, and the signal can be changed in a more unobtrusive fashion.

Using the parity coding method, the first three bits of the message 'HEY' are encoded in the following figure. Even parity is desired. The decoding process extracts the secret message by calculating and lining up the parity bits of the regions used in the encoding process. Once again, the sender and receiver can use a shared secret key as a seed in a pseudorandom number generator to produce the same set of sample regions.

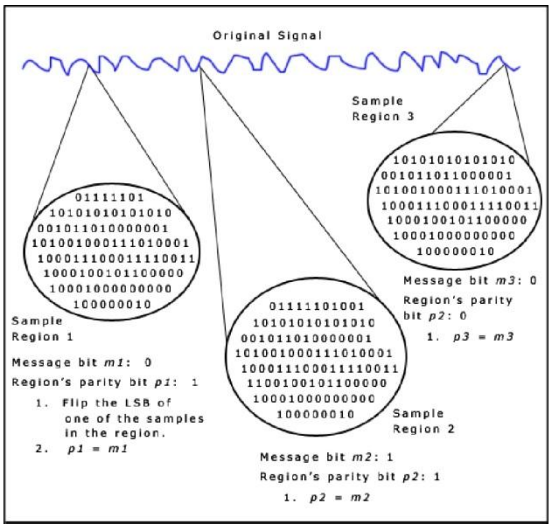


Fig.3.2. First three bits of the message 'HEY' are encoded using Parity coding method

There are two main disadvantages associated with the use of methods like LSB coding or parity coding. The human ear is very sensitive and can often detect even the slightest bit of noise introduced into a sound file, although the parity coding method does come much closer to making the introduced noise inaudible. Both methods share a second disadvantage however, in that they are not robust. If a sound file embedded with a secret message using either LSB coding or parity coding was resampled, the embedded information would be lost. Robustness can be improved somewhat by using a redundancy technique while encoding the secret message. However, redundancy techniques reduce data transmission rate significantly.

Phase Coding

Phase coding addresses the disadvantages of the noise-inducing methods of audio steganography. Phase coding relies on the fact that the phase components of sound are not as perceptible to the human ear as noise is. Rather than introducing perturbations, the technique encodes the message bits as phase shifts in the phase spectrum of a digital signal, achieving an inaudible encoding in terms of signal-to-perceived noise ratio.

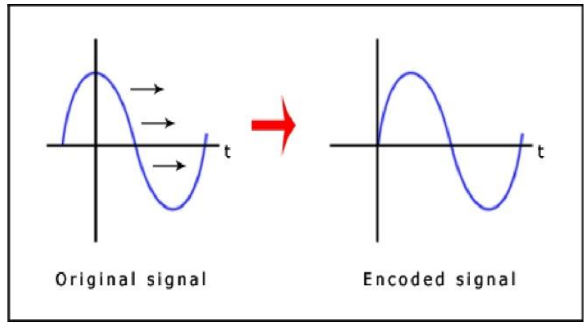


Fig.3.3. Phase Coding

To extract the secret message from the sound file, the receiver must know the segment length. The receiver can then use the DFT to get the phases and extract the information.

One disadvantage associated with phase coding is a low data transmission rate due to the fact that the secret message is encoded in the first signal segment only. This might be addressed by increasing the length of the signal segment. However, this would change phase relations between each frequency component of the segment more drastically, making the encoding easier to detect. As a result, the phase coding method is used when only a small amount of data, such as a watermark, needs to be concealed.

In a normal communication channel, it is often desirable to concentrate the information in as narrow a region of the frequency spectrum as possible in order to conserve available bandwidth and to reduce power. The basic spread spectrum technique, on the other hand, is designed to encode a stream of information by spreading the encoded data across as much of the frequency spectrum as possible. This allows the signal reception, even if there is interference on some frequencies. While there are many variations on spread spectrum communication, we concentrated on Direct Sequence Spread Spectrum encoding (DSSS). The DSSS method spreads the signal by multiplying it by a chip, a maximal length pseudorandom sequence modulated at a known rate. Since the host signals are in discrete-time format, we can use the sampling rate as the chip rate for coding. The result is that the most difficult problem in DSSS receiving, that of establishing the correct start and end of the chip quanta for phase locking purposes, is taken care of by the discrete nature of the signal.

Spread Spectrum

In the context of audio steganography, the basic spread spectrum (SS) method attempts to spread secret information across the audio signal's frequency spectrum as much as possible. This is analogous to a system using an implementation of the LSB coding that randomly spreads the message bits over the entire sound file. However, unlike LSB coding, the SS method spreads the secret message over the sound file's frequency spectrum, using a code that is independent of the actual signal. As a result, the final signal occupies a bandwidth in excess of what is actually required for transmission [6].

Two versions of SS can be used in audio steganography: the direct-sequence and frequency-hopping schemes. In direct-sequence SS, the secret message is spread out by a constant called the chip rate and then modulated with a pseudorandom signal. It is then interleaved with the cover-signal. In frequency -hopping SS, the audio file's frequency spectrum is altered so that it hops rapidly between frequencies.

Echo Hiding:

In echo hiding, information is embedded in a sound file by introducing an echo into the discrete signal. Like the spread spectrum method, it too provides advantages in that it allows for a high data transmission rate and provides superior robustness when compared to the noise inducing methods. To hide the data successfully, three parameters of the echo are varied:

Amplitude, decay rate, and offset (delay time) from the original signal. All three parameters are set below the human hearing threshold so the echo is not easily resolved. In addition, offset is varied to represent the binary message to be encoded. One offset value represents a binary one, and a second offset value represents a binary zero.

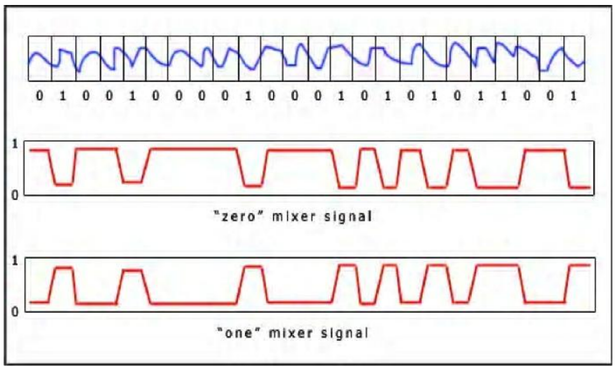


Fig.3.4. Echo Hiding

IV. Methodology

Proposed E-Cipher E & D Algorithm [5]

i. Take three key e1, e2, e3 and assign a character e1 be 'a' and e2 be 'D' and e3 be's'.

ii. Let ASCII value of e1 be 1 and e2 be 2 and e3 be 3 and take the text , add ASCII value of e1 to value of first character, and e2 to second character and e3 to third character, alternatively add the value of e1 , e2, e3 to consecutive characters.

iii. Three layers to be applied to each three consecutive letters and same to be continued thru the remaining text.

iv. After adding ASCII value of all values of given text, the resultant text is an encrypted message, and it generate a combination of 3* (256 * 256 * 256) letters encrypted coded text with 128 bit manner.

v. Transposition takes place in each character after all the process is over that is moves or change one bit either LSB or MSB, the end result is increasing security.

vi. Reverse process of the above algorithm gives the actual plain text without any error.

V. Methodologies

ISS- IHAS Sender Algorithm

Input: Audio file, Key and Original message Output: Mixed Data.

Algorithm

Step 1: Load the audio file (AF) of size 12 K.
Step 2: Input key for encryption
Step 3: Convert the audio files in the form of bytes and this byte values are represented in to bit patterns.
Step 4: Using the key, the original message is encrypted using E-Cipher algorithm.
Step 5: Split the audio file bit patterns horizontally into two halves.
Step 6: Split the Encrypted message bit patterns vertically into two halves.
Step 7: Insert the LSB bit of the vertically splitted encrypted text file (TF) into the LSB bit of the horizontally splitted audio file.
Step 8: Repeat Step 7 for the remaining bits of encrypted text file.
Step 9: If size (AF) ≥ size (TF) then
embedding can be done as explained above
else
The next higher order bit prior to previous bit position can be used
Until it is exhausted.

ISS- IHAS Algorithm - at the Receiver side:

Input: Mixed data, Key Output: Original message, audio file.

Algorithm

Step 1: Load the Stegno signal
Step 2: Extract the hidden data and audio files bit patterns from mixed data [9]
// Reverse process of step 7 of ISS-IHAS algorithm at sender side.

Step 3: Input key for decryption (as used in encryption)
Step 4: Combine the two halves of audio files bit patterns.
Step 5: Combine the two halves of encrypted messages bit pattern.
Step 6: Using Key, decrypt the original message.

VI. EVALUATION AND ANALYSIS REPORT

	Plain Text	Image	Audio	Video
Invisibility	Medium	High	High	High
Payload Capacity	Low	Low	High	High
Robustness against Statistical Attacks	Low	Medium	High	High
Robustness against Text Manipulation	Low	Medium	High	High
Variation in file size	Medium	Medium	High	Medium

Table 5.1 shows the different levels of satisfaction Level.

VII. CONCLUSION

In this paper we have introduced a robust method of imperceptible audio data hiding. This system is to provide a good, efficient method for hiding the data from hackers and sent to the destination in a safe manner. This proposed system will not change the size of the file even after encoding and also suitable for any type of audio file format. Thus we conclude that audio data hiding techniques can be used for a number of purposes other than covert communication or deniable data storage, information tracing and finger printing, tamper detection

This proposed system provides an efficient method for hiding the data from the eavesdropper. LSB data hiding technique is the simplest method for inserting data into audio signals. ISS- IHAS model is able to ensure secrecy with less complexity at the cost of same memory space as that of encrypted text and the user is able to enjoy the benefits of cryptography and steganography [7] combined together without any additional overhead. This work is more suitable for automatic control of robotic systems used in military and defense applications that can listen to a radio signal and then act accordingly as per the instructions received. By embedding the secret password in the audio signal the robot can be activated only if the predefined password matches with the incoming password that reaches the robot through audio signal. It can then start functioning as per the instructions received in the form of audio signal. More such sort of applications can be explored but confined to audio medium usage.

REFERENCES

[1] Bethany Delman,'Genetic Algorithms in Cryptography' published in web; july 2004.

[2] Darrell Whitley,'A Genetic Algorthm Tutorial', Computer Science Department, Colorado State University, Fort Collins, CO 80523

[3] Nalani N, G. Raghavendra Rao,' Cryptanalysis of Simplified Data Encryption Standard via Optimisation Heuristics;IJCSNS, Vol.6 No.1B, January 2006.

[4] Sean Simmons,'Algebric Cryptoanalysis of Simplified AES', October 2009;33,4;Proquest Science Journals Pg.305.

[5] Sujith Ravi, Kevin Knight,'Attacking Letter Substitution Ciphers with Integer Programming',Oct 2009,33,4; Proquest Science Journals Pg.321.

[6] Verma, Mauyank Dave and R.C Joshi,'Genetic Algorithm and Tabu Search Attack on the Mono Alphabetic Substitution Cipher in Adhoc Networks; Journal of COmputer Science 3(3): 134-137, 2007.

[7] William Stallings,"Cryptography and Network Security: Principles and Practice",2/3e Prentice hall , 2008.

[8] Ingemar J. Cox, Ton Kalker, Georg Pakura and Mathias Scheel. “Information Transmission and Steganography”, Springer, Vol. 3710/2005, pp. 15-29.

[9] K.Geetha , P.V. Vanthia muthu ,” International journal of Computer Science and Engineering” vol 2 No.4 Pg No: 1308-1312 , Year 2010

AUTHORS PROFILE

R. Venkateswaran received his professional degree MCA and MBA (IS) from Bharathiar University, Tamilnadu, India, He received his M.Phil from Bharathidasan University, Tamilnadu, India, and He is currently a Ph.D Scholar in the Karpagam Academy of Higher Education, Karapagam University, Tamilnadu, India, in the field of Cryptography and Network Security. Presently he is working as an Asst. Professor of Computer Applications, Nehru College of Management, Coimbatore, Tamilnadu. He is the member of CSI and other IT forums. He had published two International Journals and Presented many papers in national and International conferences, seminars and workshops. His research interests are in Cryptography and network security, information security, Software engineering,

Dr. V. Sundaram received his professional degree M.Sc. in Mathematics from the University of Madras in the year 1967 and he received his Professional Doctoral Degree Ph. D in Mathematics from the University of Madras in 1989. And He is currently working as Director, Department of Computer Applications in Karpagam College of Engineering, Tamilnadu, India, He is a research Guide for Anna university, Bharathiar university as well as Karpagam University in the field of Computer applications. He published several papers in International Journals and Conferences and also published 13 books in the area of engineering mathematics and he is the life member of ISTE and ISIAM. His research interests are in Cryptography and network security, Applied Mathematics, Discrete Mathematics, Network etc.

Image Compression using Approximate Matching and Run Length

Samir Kumar Bandyopadhyay, Tuhin Utsab Paul, Avishek Raychoudhury
Department of Computer Science and Engineering,
University of Calcutta
Kolkata-700009, India

***Abstract*— Image compression is currently a prominent topic for both military and commercial researchers. Due to rapid growth of digital media and the subsequent need for reduced storage and to transmit the image in an effective manner Image compression is needed. Image compression attempts to reduce the number of bits required to digitally represent an image while maintaining its perceived visual quality. This study concentrates on the lossless compression of image using approximate matching technique and run length encoding. The performance of this method is compared with the available jpeg compression technique over a wide number of images, showing good agreements.**

Keywords- lossless image compression; approximate matching; run length.

I. Introduction

Images may be worth a thousand words, but they generally occupy much more space in a hard disk, or bandwidth in a transmission system, than their proverbial counterpart. So, in the broad field of signal processing, a very high-activity area is the research for efficient signal representations. Efficiency, in this context, generally means to have a representation from which we can recover some approximation of the original signal, but which doesn't occupy a lot of space. Unfortunately, these are contradictory requirements; in order to have better pictures, we usually need more bits.

The signals which we want to store or transmit are normally physical things like sounds or images, which are really continuous functions of time or space. Of course, in order to use digital computers to work on them, we must digitize those signals. This is normally accomplished by sampling (measuring its instantaneous value from time to time) and finely quantizing the signal (assigning a discrete value to the measurement) [1]. This procedure will produce long series of numbers. For all purposes of this article, from here on we will proceed as if these sequences were the original signals which need to be stored or transmitted, and the ones we will eventually want to recover. After all, we can consider that from this digitized representation we can recover the true (physical) signal, as long as human eyes or ears are concerned. This is what happens, for example, when we play an audio CD. In our case, we will focus mainly on image representations, so the corresponding example would be the display of a picture in a computer monitor. However, the discussion in this paper, and especially the theory developed here, apply equally well to a more general class of signals.

There are many applications requiring image compression, such as multimedia, internet, satellite imaging, remote sensing, and preservation of art work, etc. Decades of research in this area has produced a number of image compression algorithms. Most of the effort expended over the past decades on image compression has been directed towards the application and analysis of different coding techniques to compress the image data. Here in this paper also, we have proposed a two step encoding technique that transform the image data to a stream of integer values. The number of values generated by this encoding technique is much less than the original image data. The main philosophy of this encoding technique is based on the intrinsic property of most images, that similar patterns are present in close locality of images.

The coding technique makes use of this philosophy and uses an approximate matching technique along with the concept of run length to encode the image data into a stream of integer data. Experimental results over a large number of images have shown good amount of compression of image size.

II. Related Works

Image compression may be lossy or lossless. Lossless compression is preferred for archival purposes and often for medical imaging, technical drawings, clip art, or comics. This is because lossy compression methods, especially when used at low bit rates, introduce compression artifacts. Lossy methods are especially suitable for natural images such as photographs in applications where minor (sometimes imperceptible) loss of fidelity is acceptable to achieve a substantial reduction in bit rate. The lossy compression that produces imperceptible differences may be called visually lossless.

Methods for lossless image compression are:

- Run-length encoding – used as default method in PCX and as one of possible in BMP, TGA, TIFF
- DPCM and Predictive Coding
- Entropy encoding
- Adaptive dictionary algorithms such as LZW – used in GIF and TIFF
- Deflation – used in PNG, MNG, and TIFF
- Chain codes

Run-length encoding (RLE) is a very simple form of data compression in which runs of data (that is, sequences in which the same data value occurs in many consecutive data elements) are stored as a single data value and count, rather than as the original run. This is most useful on data that contains many such runs: for example, simple graphic images such as icons, line drawings, and animations. It is not useful with files that don't have many runs as it could greatly increase the file size.

DPCM or differential pulse-code modulation is a signal encoder that uses the baseline of PCM but adds some functionalities based on the prediction of the samples of the signal. The input can be an analog signal or a digital signal.

Entropy encoding is a lossless data compression scheme that is independent of the specific characteristics of the medium.

One of the main types of entropy coding creates and assigns a unique prefix-free code to each unique symbol that occurs in the input. These entropy encoders then compress data by replacing each fixed-length input symbol by the corresponding variable-length prefix-free output codeword. The length of each codeword is approximately proportional to the negative logarithm of the probability. Therefore, the most common symbols use the shortest codes.

Lempel–Ziv–Welch (LZW) is a universal lossless data compression algorithm created by Abraham Lempel, Jacob Ziv, and Terry Welch. It was published by Welch in 1984 as an improved implementation of the LZ78 algorithm published by Lempel and Ziv in 1978. The algorithm is simple to implement, and has the potential for very high throughput in hardware implementations.

Deflate is a lossless data compression algorithm that uses a combination of the LZ77 algorithm and Huffman coding. It was originally defined by Phil Katz for version 2 of his PKZIP archiving tool, and was later specified in RFC 1951.

A chain code is a lossless compression algorithm for monochrome images. The basic principle of chain codes is to separately encode each connected component, or "blot", in the image. For each such region, a point on the boundary is selected and its coordinates are transmitted. The encoder then moves along the boundary of the image and, at each step, transmits a symbol representing the direction of this movement. This continues until the encoder returns to the starting position, at which point the blot has been completely described, and encoding continues with the next blot in the image.

III. OUR WORK

The main philosophy behind selecting approximate matching technique along with run length encoding technique is based on the intrinsic property of most images, that they have similar patterns in a localized area of image, more specifically the adjacent pixels row differ in very less number of pixels. This property of image is exploited to design a very effective image compression technique. Testing on a wide variety of images has provided satisfactory results. The technique used in this compression methodology is described in this section.

We consider approximate matching algorithm and run length for our image compression. The approximate matching algorithm does a comparison between two strings of equal length and represent the second string with respect to the first only with the information of the literal position where the string mismatches.

Replace. This operation is expressed as (p; char) which means replacing the character at position p by character char.

Let C denote “copy", and R denote “replace” then the following are two ways to convert the string “11010001011101010" to “11010001111001010" (0,1 are stored in ASCII)via different edit operation sequences:

C C C C C C C C R C C R C C C C C
1 1 0 1 0 0 0 1 0 1 1 1 0 1 0 1 0
1 1 0 1 0 0 0 1 1 1 1 0 0 1 0 1 0

A list of edit operations that transform a string u to another string v is called an EditTranscription of the two strings [9]. This will be represented by an edit operation sequence (u; v) that orderly lists the edit operations. For example, the edit operation sequence of the edit transcription in the above example is (\11010001011101010",\11010001111001010") = (9; 1),(12,0);

Approximate matching method. In this case, the string \11010001011101010" can be encoded as f(17; 2)= (9; 1),(12,0), storing the ASCII characters require 136 bit or 17 byte where as storing 4 characters will require 4 byte. Thus a compression of approximate 76.4% is achieved. This technique is very useful in image compression because of the inherent property of an image because two consecutive rows of an image has almost same string of pixel values. Only a few pixel varies. Experimental results prove this hypothesis.

Apart from the concept of approximate matching method, the concept of run length is also used because using run length a row of image can be represented using much less literals than the original.

Run-length Encoding, or RLE is a technique used to reduce the size of a repeating string of characters. This repeating string is called a run, typically RLE encodes a run of symbols into two bytes , a count and a symbol. RLE can compress any type of data regardless of its information content, but the content of data to be compressed affects the compression ratio. Consider a character run of 15 'A' characters which normally would require 15 bytes to store :

AAAAAAAAAAAAAAA is stored as 15A

With RLE, this would only require two bytes to store, the count (15) is stored as the first byte and the symbol (A) as the second byte.

In this compression technique, we have used the approximate matching method in unison with run length. Starting from the left uppermost row of image, every three rows are considered at a time. Of these, the middle row is represented using run length, and the row above and below it are matched with the middle row using approximate matching method. This method is continued iteratively until the whole image is scanned and compressed.

The algorithms designed as per our technique are as follows:

A. COMPRESS (Source Raw Image file)

This is the main algorithm for compression. This algorithm will be used to compress the data part of the Source Image File.

Output: It will output the Compressed-Image file.

Input: This function will take Source Image file as input.

1. Read the Source Image file as input. Obtain it's size (say r*c).Store the data part(pixel values) of the image in an array A of the same size.
2. Quantize the color palate of the image, i.e array A with quantization factor 17.
3. If r is not divisible by 3, then duplicate the last row 1 or 2 times at the bottom of A such that the number of rows become divisible by 3. Reset r with the corresponding new size of A.
4. Take a blank array say 'Compress' of size n*2(n is a positive integer).Starting with the 1st row, choose consecutive 3 rows at a time and perform the following operations(say, we have chosen row number k-1,k and k+1) in each iterations:
   a. For each column in the array A, if any mismatch is found in the row k-1 and k, the corresponding column number and the value at that corresponding column with row number k-1 in array A, is stored in array Compress. For every mismatch, those two values are stored in a single row of array Compress.
   b. For row number k, the corresponding value (starting from the 1st column) and its runlength (Number of consecutive pixels with same pixel value) for kth row is stored in array Compress. Every set of value and it's runlength is stored in a single row of array Compress.
   c. For each column in the array A, if any mismatch is found in the row k and k+1, the corresponding column number and the value at that corresponding column with row number k+1 is stored in array Compress. For every mismatch, those two values are stored in a single row of array Compress.
5. Repeat Step 4 until all the rows are compressed. A marker should be used to distinguish between the encrypted versions of each row. Store also the value of 'r' and 'c' in Compress.
6. Array Compress now constitutes the compressed data part of the corresponding Source Image File.

B. DECOMPRESS (Compressed-Image file)

This is the main algorithm for decompression or decoding the image. This algorithm will be used to decompress the data part of the Source Image File i.e. the image from the 'Compress' array.

Output: It will output the Decompressed or Decoded Image file.

Input: This function will take the Compressed-Image file ('Compress' array) as input.

1. Read the Compress array. Obtain the size of the image (say r*c) from the array. Take a blank array say 'Rec' of the same size for reconstruction of the data part of the image.
2. Starting from the 1st row, consider the compressed values of consecutive 3 rows from Temp and perform the following operations(say, we have chosen row number k-1,k and k+1) in each iterations:
   a. Firstly, construct the kth row of Rec array with the corresponding positional value and runlength value in the Compress array, by putting the same positional value in runlength number of consecutive places in the same row.
   b. Then, construct the (k-1)th row. In the corresponding Compress array for this particular row for each column if an entry for column number 'v' is not present, then Rec[(k-1),v]=Rec[k,v]. Else, if Compress[i,1]=v then, Rec[(k-1),v]= Compress[i,2].
   c. Then, construct the (k+1)th row. In the corresponding Compress array for this particular row for each column if an entry for column number 'v' is not present, then Rec[(k+1),v]=Rec[k,v]. Else, if Compress [i,1]=v then, Rec[(k+1),v]= Compress [i,2].
3. Step 2 is repeated until the full Rec array is filled.
4. Rec array is stored as the Decompressed Image File.

IV. RESULT AND DISCUSSION

A. Complexity analysis of the stated algorithm

Let the size of the image be r*c. Then, at the time of Compression, 3 rows are considered at a time and for each compression of rows 'c' number of columns are read. At this process 3 rows are compressed at a time taking 3*c number of comparisons. So, for compression of the whole image, total number of compression required is $\frac{r}{3}$ *3*c=r*c ,that is O(r*c).So, for an image of size n*n, the time complexity of the compression algorithm is O(n2).

In the receiver end, the Compress array is read and the Rec array is reconstructed which also takes number of comparisons=r*c ,that is O(r*c).So, for an image of size nxn, the time complexity of the de-compression algorithm is O(n2).

B. Test Results

Before Compression (For each image) :
Size : 300 x 280 = 84000 [Row x Col]
Size in bytes : 1,68000 byte = 168kb
After Compression (For each image) :

Figure 1. Nature

Size in bytes : 51348 byte = 102.69 kb

Compression percentage : 38.87 %

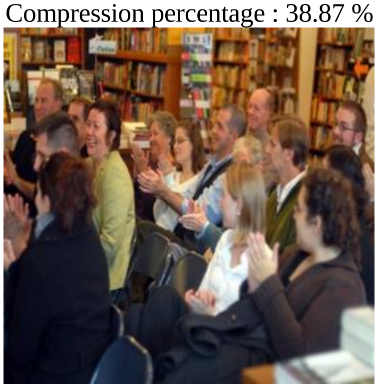


Figure 2. Library

Size in bytes : 55234 byte = 110.64 kb

Compression percentage : 34.14 %

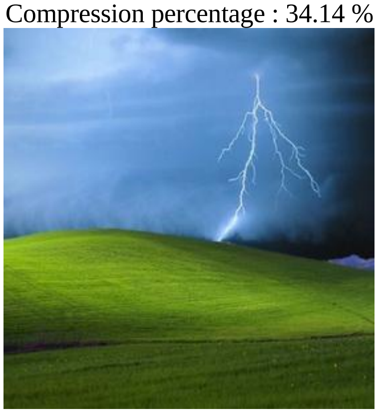


Figure 3. Landscape

Size in bytes : 28790 byte = 57.58 kb

Compression percentage : 65.73 %

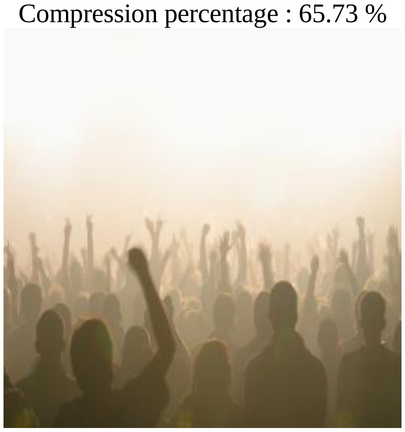


Figure 4. Crowd

Size in bytes : 13616 byte = 27.23 kb

Compression percentage : 83.79 %

Figure 5. Tom_Jerry

Size in bytes : 35504 byte = 71.00 kb

Compression percentage : 57.73 %

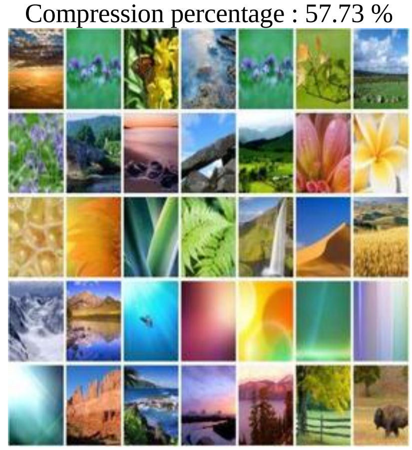


Figure 6. Thumbnail

Size in bytes : 76680 byte = 153.36 kb

Compression percentage : 8.71 %

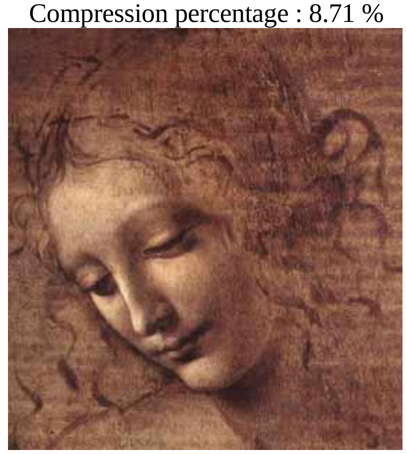


Figure 7. Model_face

Size in bytes : 63094 byte = 126.18 kb

Compression percentage : 24.89 %

C. Conclusion

The algorithm proposed here is for lossless image compression ass it is evident from the algorithm, that the exact image data (pixel values) are extracted from the compressed data stream without any loss. This is possible because the compression algorithm does not ignores or discards any original pixel value. Moreover the techniques such as approximate matching and run length encoding technique are intrinsically lossless.

This compression technique proves to be highly effective for images with large similar locality of pixel lay out. This technique will find extensive use in medical imaging sector because of its lossless characteristics and the medical images has large area of similar pixel layout pattern, like in X – ray images large area are black.

REFERENCES

[1] P. S. R. Diniz, E. A. B. da Silva, and S. L. Netto, Digital Signal Processing: System Analysis and Design. Cambridge University Press, 2002.

[2] J. M. Shapiro, “Embedded Image Coding Using Zero-Trees of Wavelet Coefficients”, IEEE Transactions on Signal Processing, vol. 41, pp. 3445–3462, December 1993.

[3] A. Said and W. A. Pearlman, “A New, Fast and Efficient Image Codec Based on Set Partitioning in Hierarchical Trees”, IEEE Transactions on Circuits and Systems for Video Technology, vol. 6, no. 3, pp. 243– 250, June 1996.

[4] D. Taubman, “High Performance Scalable Image Compression with EBCOT”, IEEE Transactions on Image Processing, vol. 9, no. 7, July 2000.

[5] Tse-Hua Lan and A. H. Tewfik, “Multigrid Embedding (MGE) Image Coding”, Proceedings of the 1999 International Conference on Image Processing, Kobe.

[6] Civarella and Moffat. Lossless image compression using pixel reordering. Proceedings of twenty seventh Australian Computer Science conference,pp 125-132,2004.

[7] K.Veeraswamy, S.Srinivaskumar and B.N.Chatterji. Lossless image compression using topological pixel re-ordering. IET international conference, India, pp 218-222,2006.

[8] Memon and Shende. An analysis of some scanning techniques for lossless image coding. IEEE Transactions on Image Processing,9 (11),pp 1837-1848,2000.

[9] Memon and Wu X. Recent developments in context based predictive techniques for lossless image compression. The computer Journal,40,pp 127-136,1997.

[10] K.Sayood. Introduction of data compression. Acdemic press, 2nd edition,2000.

[11] D.Salomon. Data Compression. Springer,2nd edition, 2000.

AUTHORS PROFILE

Dr. Samir K. Bandyopadhyay, B.E., M.Tech., Ph. D (Computer Science & Engineering), C.Engg., D.Engg., FIE, FIETE, currently, Professor of Computer Science & Engineering, University of Calcutta, visiting Faculty Dept. of Comp. Sc., Southern Illinois University, USA, MIT, California Institute of Technology, etc. His research interests include Bio-medical Engg, Image processing, Pattern Recognition, Graph Theory, Software Engg.,etc. He has 25 Years of experience at the Postgraduate and under-graduate Teaching & Research experience in the University of Calcutta. He has already got several Academic Distinctions in Degree level/Recognition/Awards from various prestigious Institutes and Organizations. He has published more than 300 Research papers in International & Indian Journals and 5 leading text books for Computer Science and Engineering. He has visited round the globe for academic and research purposes.

Tuhin Utsab Paul received his Bachelors degree in Computer science in 2008 and Masters degree in Computer and Information Science in 2010, both from the University of Calcutta. He is currently doing his M.Tech course in Computer science and engineering from the University of Calcutta. His research interest include Image processing, pattern recognisation, image steganography and image cryptography. He has published quite a few Research papers in International & Indian Journals and conferences. He is an IEEE member since 2009.

Avishek Raychoudhury received his Bachelors degree in Computer science in 2008 and Masters degree in Computer and Information Science in 2010, both from the University of Calcutta. He is currently doing his M.Tech course in Computer science and engineering from the University of Calcutta. His research interest include Image processing, image steganography, graph theory and its applications. He has published quite a few Research papers in International & Indian Journals and conferences.

Interactive Intranet Portal for effective Management in Tertiary Institution

Idogho O. Philipa
Department of Office Technology and Management
Auchi Polytechnic, Auchi
Edo State, Nigeria

Akpado Kenneth
Department of Electronics and Computer Engineering
Nnamdi Azikiwe University
Anambra State Nigeria

James Agajo
Electrical/ Electronics Dept.
Auchi Polytechnic, Auchi
Edo State, Nigeria

Abstract— **Interactive Intranet Portal for effective management in Tertiary Institution is an enhanced and interactive method of managing and processing key issues in Tertiary Institution, Problems of result processing, tuition fee payment, library resources management are analyzed in this work. An interface was generated to handle this problem; the software is an interactive one. Several modules are involved in the paper, like: LIBRARY CONSOLE, ADMIN, STAFF, COURSE REGISTRATION, CHECKING OF RESULTS and E-NEWS modules. The server computer shall run the portal as well as OPEN SOURCE Apache Web Server, MySQL Community Edition RDBMS and PHP engine and shall be accessible by client computers on the intranet via thin-client browser such as Microsoft Internet Explorer or Mozilla Firefox. It shall be accessible through a well-secured authentication system. This Project will be developed using OPEN SOURCE technologies such XAMMP developed from WAMP (Windows Apache MySQL and PHP)**

Keywords- Portal; Database; webA; MYSQL; Intranet; Admin.

I. Introduction

Interactive intranet portal for effective management in tertiary institution seeks to address the Problems arising from result processing, tuition fee payment, library resources management are analyzed in this work. An interface was generated to handle this problem, the software is an interactive one. An intranet is a network inside an organization that uses internet technologies (such as web browsers and servers, TCP/IP networks protocols, HTML hyper media document publishing and databases, etc) to provide an intranet-like environment with the organization for information sharing, communications, collaboration, and the support of business processes. An intranet is protected by security measures such as passwords, encryption, and firewalls, and thus can be accessed by authorized users through the intranet. Secure intranets are now the fastest-growing segment of the internet because they are less expensive to build and manage than private networks based on proprietary protocols.

Internets appeared in the mid-1990s and were perceived as the answer to the need for the integration of existing information systems into organizations. Despite the fact that has been extensive research regarding implementation, development processes, policies, standardization vs. creativity and so forth, the potentiality of intranets has not been fully exploited. Intranets offer many advantages in the form of working networks that support and enables empowered employees to participate in the development of the organization, to enable the measurement of essential functions and to monitor industries conditions and find suitable functions that support doing work.[1]

II. Background

On our Institution like many other universities there is an intense need for communication and co-operation between the administrational staff and department. This is because most of the department resources like student course registration, Result management, staff management and student management have to be managed partly by one or the other group, different resources for different reasons.

III. Objectives

The main objective of this project is to develop on intranet portal software that will adequately manage records

A. Benefits

1. Enable the library to keep inventory of all books
2. Keep track of borrowed books
3. Save information of all student id cards
4. Show student detail and books borrowed by these students
5. Provide easy collation of students information for clearance
6. Connect with oxford university library for additional materials
7. Web based access for e-books, papers, journals and students projects which promote Auchi Polytectnic worldwide via the internet.
8. Access journals and papers from Oxford University press.
9. A computerized easy and fast clearance process in the library

B. Justification

- Need for efficient, effective and adequate management of students records[2]

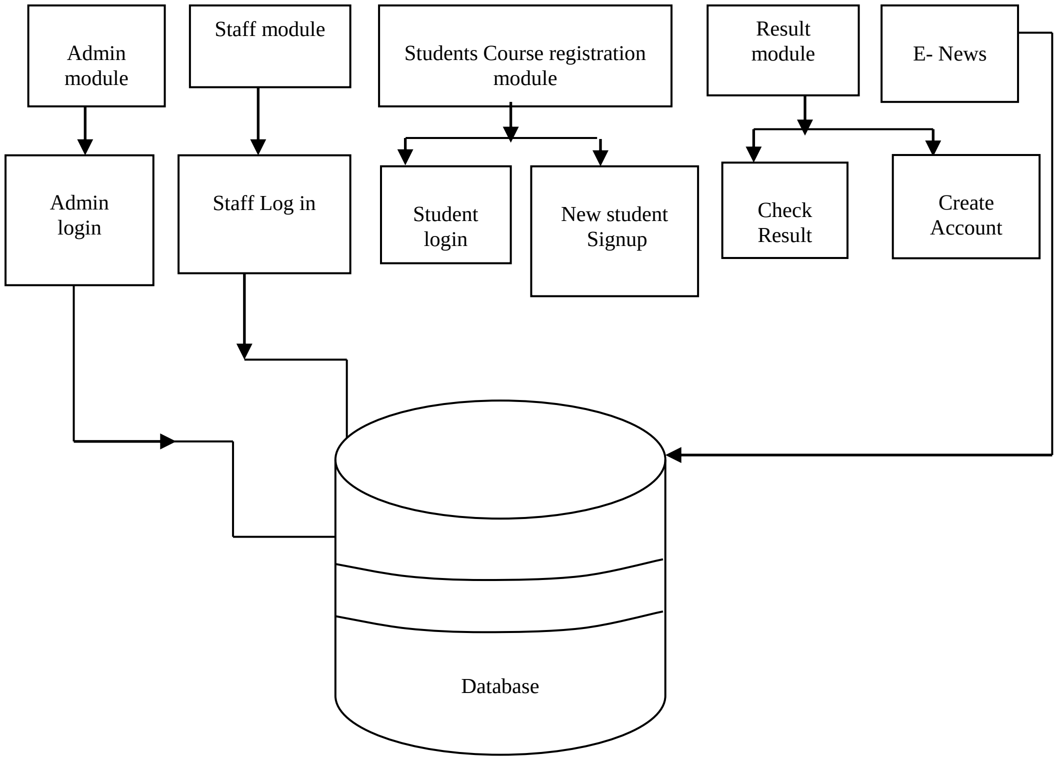


Figure 1 BLOCK DIAGRAM OF PORTAL

- Need for adequate protection and security of vital information

Providing academics the ability to manage and communicate more effectively with students data.

- Helping academics spend less time on processing students data
- Need for easy and past means of information dissemination within the department

IV. METHODOLOGY

Being a web-based portal, first the web development is determined using an open sources platform that will be more flexible and have lower cost due to free licensing.

Dreamweaver: This is a unique tool expressly designed for the development and optimization of web pages. All the coding in this project is done in the code view of the Dreamweaver.

Hypertext preprocessor (PHP): PHP is the web development language written by and for web developers .PHP is a server-side scripting language, which can be embedded in the HTML. In this project PHP is used to write all the forms script codes which made the software very interactive and more dynamic.

MYSQL: This is an open source, SQL Relational Database Management System (RDBMS) that is free for many uses. MYSQL is used for database management in this project.

V. CORPORATE PORTAL DEFINITIONS

Figure 1 show a block diagram describing various stages in the portal, a portal was referred to as search engine, whose main goal was to facilitate access to information contained in documents spread throughout internet. Initially, search engines enabled internet users to locate documents with the use of Boolean operators or associative links between web pages. To reduce even more the searching time and to help inexperienced users, some search engines have included categories, that is, they started to filter sites and sports, metrology, tourism, finance, news, culture etc. The succeeding steps were the integration of other functions, such as virtual communities and real time chats; the ability to personalize search engine interfaces (my yahoo, my Excite, etc); and to access specialized and commercial contents. This new concept of search engine is now called a portal.

VI. MAJOR CHARACTERISTICS OF A CORPORATE PORTAL

Since corporate portals integrate some well-know technologies, such as intelligence business tools, document management, office automation, groupware, data warehouse, and intranet, some suppliers of products on these areas have also positioned themselves as corporate portal vendors. At the same time, small companies have viewed the great market opportunity of corporate portals and have announced new portal products. Besides, some big computers companies have established technical and/or commercial alliance to provide joint solutions and to suit specific needs of their customers. Therefore, the solution of a particular corporate portal, amongst all products available on the market today, is not an easy task. These are some of the characteristics of corporate portal.

- The ability to manage the information life cycle, establishing storage hierarchical levels and discarding necessary

❖ The ability to locate experts in the organization, in accordance with the type of knowledge demanded for a particular task

❖ The ability to satisfy the information needs of all types of corporate users.

❖ The possibility of information exchange among customers, employees, suppliers and resellers, providing an information infrastructure suitable for electronic commerce.

Conclusively, corporate portals, whose "ancestors" are the decision support and the management information systems, are the next step into modern design of user interfaces to corporate information. Adapting the enterprise environment to suit users' needs and optimize the interaction, distribution and management of internal and external information resources the corporate portal allow users to access corporate information in an easier and customized way, resulting, the theoretically, in reduced costs, increased, productivity and competitiveness.

VII. System Design Approach

A. Top-Down Design

In top-down design, the system is designed in such a way that all necessary steps needed for the realization of the design will be followed promptly starting from the known to the unknown. The first stage is the approval of the project topic/title followed by the necessary information/data gathering required for the design stage. Next is the designing stage that is the main part of the work followed by the implementation of the design work where all the required components are being applied. The testing stage where the implemented works are tested in order to determine the output of the system.[3]

B. Benefits

- To save time and reduce the stress involved in student's result processing;
- To facilitate student's result processing activities with the aid of a program to give rise to effective production of student's results with fewer human errors;
- To handle large volumes of results processing;
- To aid the management in the providing results data in time.
- To design a computerized result processing system that will minimize the amount of manual input needed in student result processing.

Such a system would help the polytechnic system in becoming more computer-friendly and technologically-inclined as staff and students could make use of it and benefit more from its numerous objectives and advantages.

Figure 2 and figure 3 shows Flow chart of student course registration Module and Flow chart of Admin module. The flow diagram represents the stages and sequence this world pass through.

VIII. Structural Analysis

The first point to mention is the users of the system. Different user can use the portal. The system will be analyzed in the following [4]

❖ **Admin Module**: The admin module has interface that allow the administrator to login into the system and have over roll control of the system. The Admin has to create login ID and password

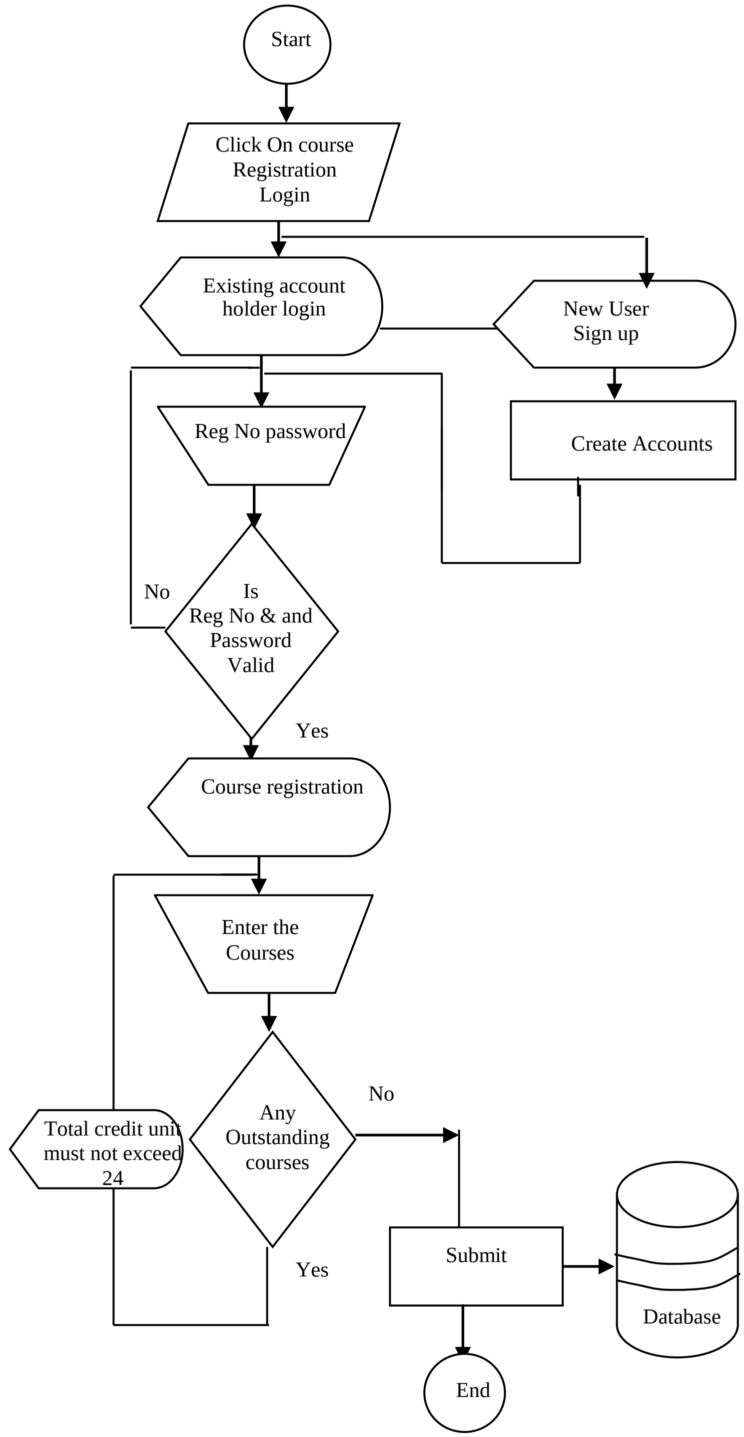


Figure 2. Flow chart of student course registration Module

to the authorized staffs. The admin decides on certain privileges like update, delete, and edit to only authorized persons. The logic flow in Admin module is represented below.

IX. Student Course Registration Module

This module has interface that allow students to register their courses. A new student who has not registered at all has to go through student account signup and obtain user name and password before being able to enroll to the program and select courses. In the signup step, the student is asked to fill his profile information such as his first name, last name, middle name, registration number, and gender, date of birth, email address, permanent address, phone number and educational level. The student has to login with the user ID and password obtained from signup account. The logic flow in the student course registration is represented below. [6]

X. Result Module

This module has interface that allow students to login into

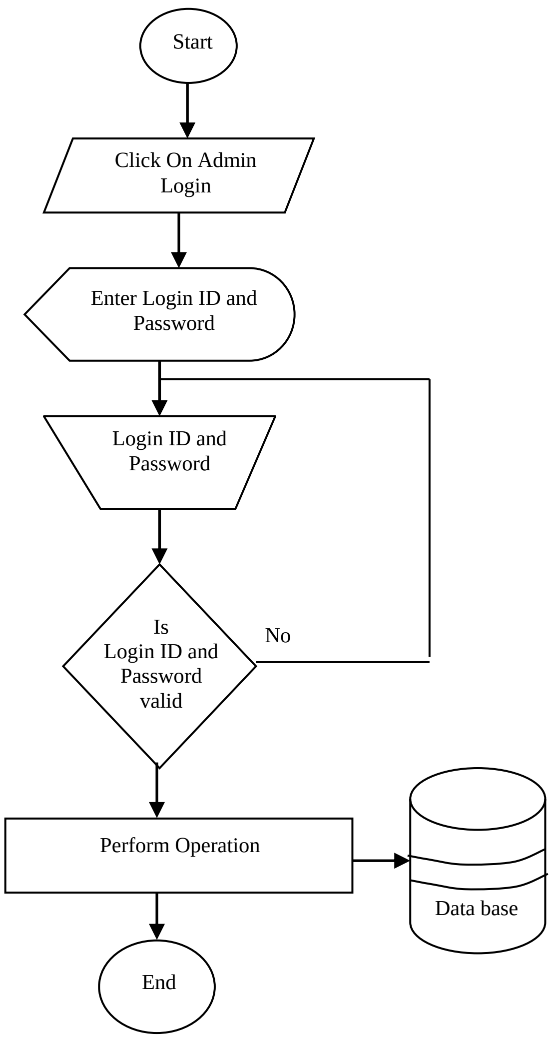


Figure 3 Flow chart of Admin module

the system, using their registration number and password. New users have to signup an account that will enable them to log in and access their result. The system uses the registration number to query the database in order to retrieve the result pointing to the registration number. The logic flow in the result module is represented below.

XI. Database Design

The database is the long-term memory of the web database application. Relational database management system (RDBMS) was considered during the design of this database. The database structures have tables and their relationships are defined. The databases have eight (8) tables and they are related with primary and foreign keys..[7]

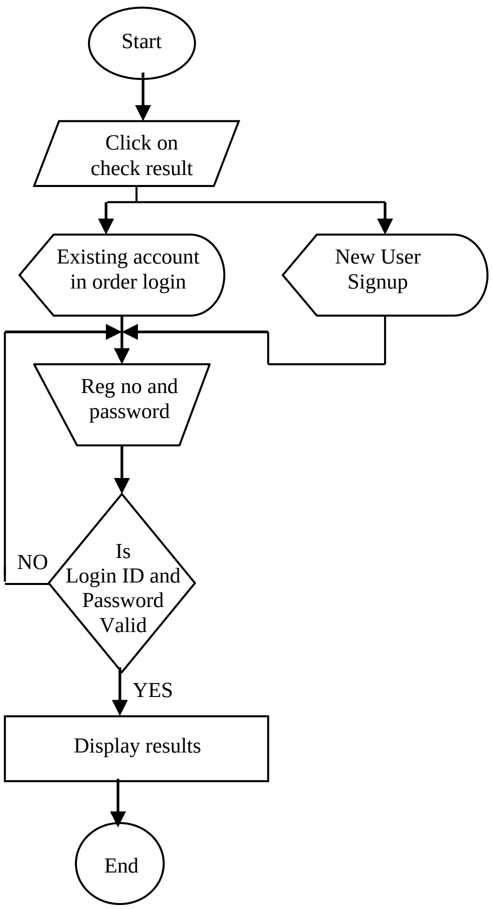


Figure 4 Flow chart of Result Modules module

1) *Course table*
2) *Course level table*
3) *Exam table*
4) *Student table*
5) *Staff table*
6) *Semester table*
7) *Session table*
8) *Year table*
9) *Result table*

XII. System Implementation

The system implementation was done using open source software XAMMP developed from WAMP (Window Apache, MYSQL and PHP) and integrating it with macromedia Dreamweaver installed on computer. The following and done for the implementation.[8],[9]

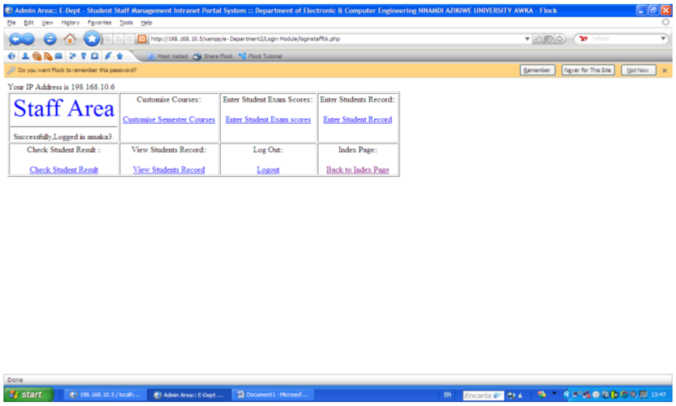


Figure 5 Proposed interface for Auchi library Management

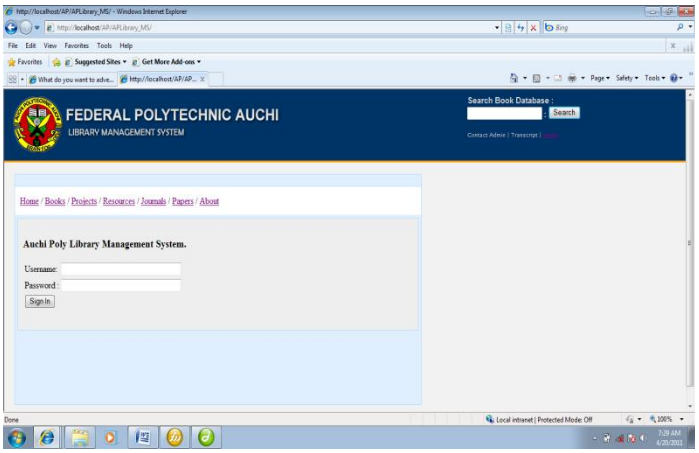


Figure 6 (SCREEN SHOT) APPLICANT STATISTICS 1

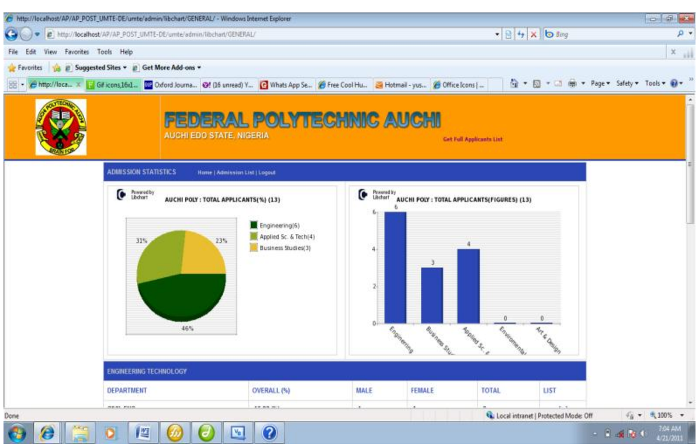


Figure 7(a) (SCREEN SHOT) APPLICANT STATISTICS 2

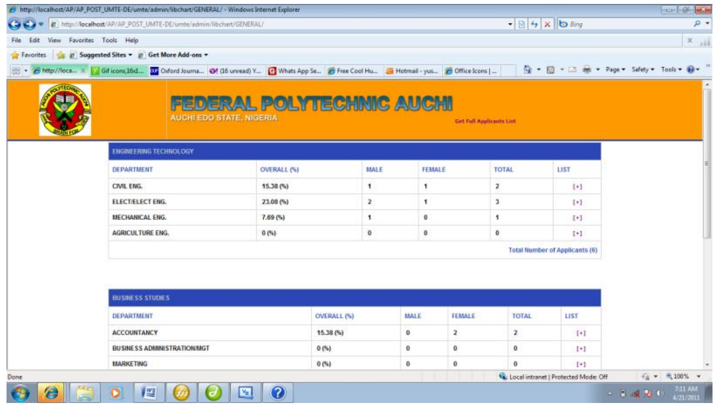


Figure. 7(b) (SCREEN SHOT) APPLICANT STATISTICS 3

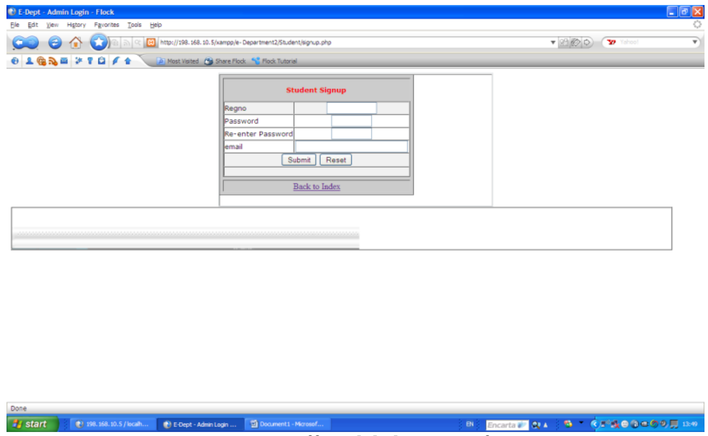


Figure. 8 Staff module login interface

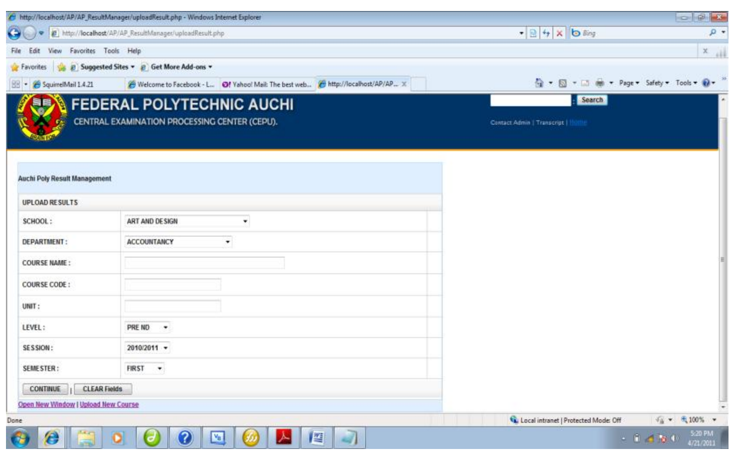


Figure 9 Student course Registration interface

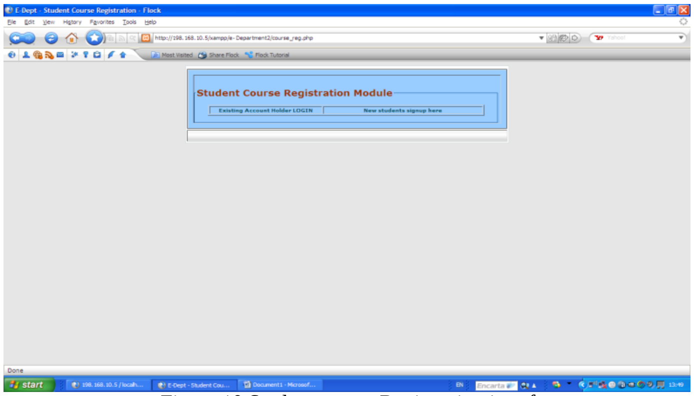


Figure 10 Student course Registration interface

1) *The modules image was created using macromedia flash 8.0*

2) *The admin module was created by following the logic flow diagram using dream weaver 8.0*

3) *The staff module was also created as in figure 8*

4) *Student course registration / check result modules was also created*

5) *E-news module was created and linked to Admin module*

Admin login was created and the login ID and password is also created and store in the database

Finally common database was designed for all the modules

XIII. Sub Modules Implementation

A. Admin Module

The following are done on implementation Admin Area was created and it has the following features

1) *Customize semester courses*
2) *Enter student Exam scores*
3) *Enter student Record*
4) *Check student result*
5) *View students record*
6) *Logout*
7) *Back to Index page*
8) *Assign login details to staff*

B. Staff Module

The following are done on Implementation

1) *Obtain login ID and password from admin*

2) *Staff Area was created and it has the following features , see figure 8*

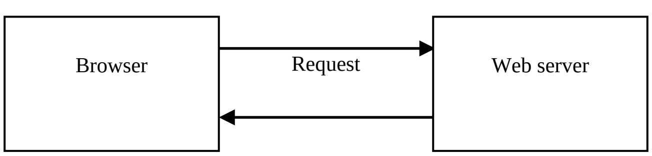


Figure 11 Student Sign up Block interface [13,14]

a) *Customize semester courses*
b) *Enter student Exam scores*
c) *Enter student Record*
d) *Check student result*
e) *View students record*
f) *Logout*
g) *Back to Index page*

XIV. Auchi Polytectnic Web Based Admission Processing Center

The system will help speed up admission processing within 24hr. The proposed system will automatically categorize every applicant within their various choices and arrange them according to their performance (Jamb Score) in an ascending order. More Explanation can be provided if requested. [8]

A. Student course registration

The following are done on implementation

1) *Student course registration as shown in figure. 10 have two sub modules was created which account is stored in the signup database table.*

a) *Existing Account Holder*

2) *Existing account holder login was also created*

3) *New student signup here was created and the data account is stored in the signup database table.*

4) *Existing account holder login was also created.*

XV. Database Implementation

Implementation of the database is carried out on MYSQL 5.0 DB sever engine. The data retrieval is analyzed in the implementation phase. The basic operation of a web server is in fig 13 and 14.

A communication link is required between them. A web browser makes a request of the server. The servers send back a response. The web database implementation in this project follow a general web database structure as shown.

The basic web database architecture consists of the web browser, webserver, scripting engine and database server

The following stages are involved in data retrieval from the database

1) *A user's web browser issues an HTTP request for a particular web page*

2) *The web server receives the request for the page, retrieves the files, and passes it to the PHP engine for processing*

3) *The PHP engine begins parsing the script. Inside the script is command to connect to the database and execute a query. PHP opens a connection to the MYSQL server and sends on the appropriate query.*

4) *The MYSQL server receives the database query, processes it and sends the results back to the PHP engine the database as shown below.*

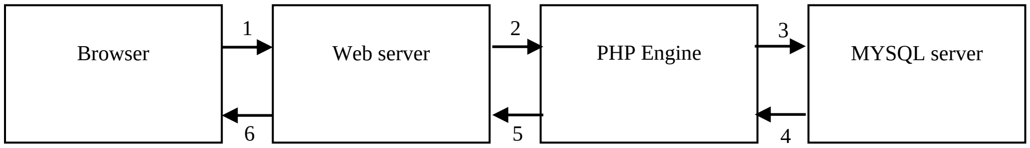


Figure 12 Student Enhanced Sign up Block interface [15,16]

A. *Database entity relationship model*

This is the relationship model in MYSQL database as shown in the figure 15

a) Creating Database table

All the database tables in my design was created in MYSQL by writing the SQL requires save in the root of my web server called SQL.Main. Proper SQL command was used to insert it into

XVI. SYSTEM TESTING AND EVALUATION

A. *Test Plan*

In testing this work, one should adopt, first the bottom-up approach to test the various modules before finally testing the complete system with the control program as was done in this work.

Furthermore, ones main concerned was the fact that the system should meets its functional and requirements.

B. *Testing the Admin Module Interface*

The module was really tested with the xampp server user Interface in figure 16. The admin was tested with the valid login ID and password that allow the admin into database and the admin area was display on the interface.

C. *Testing the Staff Module Interface*

This module was really tested with the xampp server user interface. The login ID and password obtained from admin was used to test the staff module and it was successful

D. *Testing the Student Course Registration Module Interface*

This module was really tested with the xampp server user interface. The module was tested with the login ID and password obtained by student signup account. It interacted with the database

E. *Testing the Check Result Module Interface*

This module was really tested with the xampp server user interface.

The module was also tested with the login and password obtained from student check result signup account. The testing was successful.

F. *Software Testing and Debugging*

The program development was carried out in modules using dreamweaver8.0 connected to xampp server containing Apache, MYSQL and PHP. The first task was to write all the PHP codes in Dreamweaver and save it in root of my server. Next was to check if the Apache server is running which acts as the web server and if it connected to MYSQL database.

Finally, all the modules were linked into the database to check if they are all running.

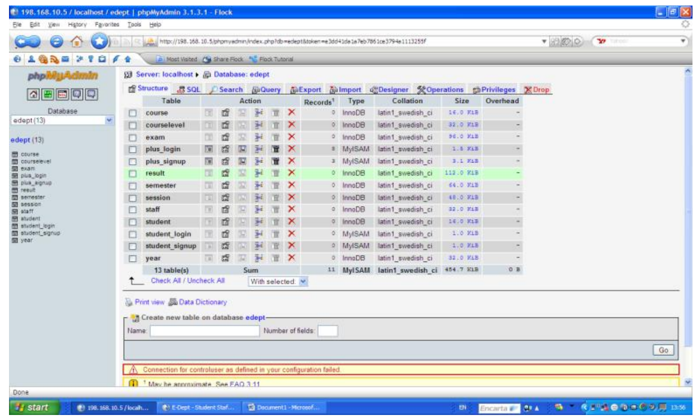
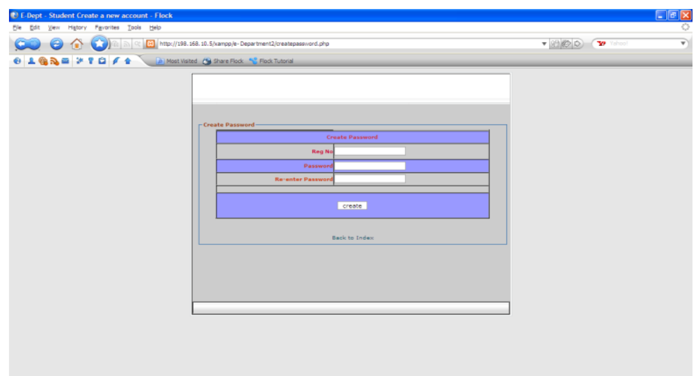


Figure 13 Student Sign up interface



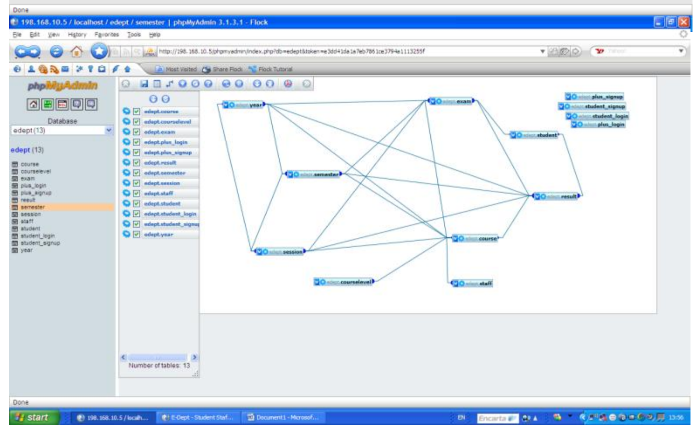


Figure. 14 :Client/Server relationship between and server

Figure. 15 Database entity relationship model interface

XVII. ACTUAL VS EXPECTED RESULTS

During random testing of all the modules some errors came up and I discovered that it was not inserting data into the database.

TABLE 1. ACTUAL RESULT VS. EXPECTED RESULT

ACTUAL RESULT	EXPECTED RESULT
It does not insert data into the database.	Expected to insert data into the database.
Could not execute SQL queries.	It is supposed to execute SQL queries.

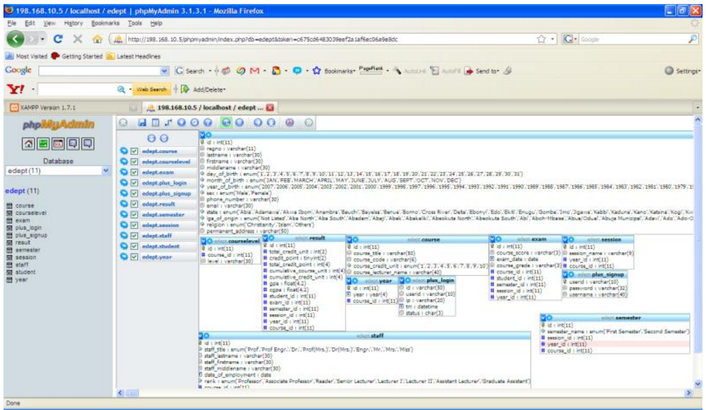


Figure. 16 Database tables interface

XVIII. PERFORMANCE EVALUATION

The performance of the Results did not meet the actual design of this project. It is expected that the project can be continued upon after the defense to meet the required design of this project.

The proliferation of digital information resources and electronic databases challenges libraries and demands that libraries develop new mechanisms to facilitate and better inform user selection of electronic databases and search tools.[11,12]

XIX. CONCLUSION

The fundamental idea of intranet portal-based designs is to reduce the amount time spent in data processing. Companies and organizations with intranet portal have attracted much stock market investors interest because portals are viewed as able to command a large audience.

Finally, intranet portal bring vast information and services resources available from many sources to many users within the same organization in an effective manner.

REFERENCES

[1] Horton Jr., F., "Infotrends", Prentice Hall 2nd edition, pp. 185-191, 1996.

[2] Oliveira, D., São Paulo: Atlas, IEICE Trans.Commun., pp 123-144, 1996.

[3] Cronin, B., & Davenport, E., "Elements of Information management". New Jersey: Scarecrow Press, 1991.

[4] Taylor, A., & Farrell, S., "Information management in context". Aslib proceedings, 44(9), pp. 319-322, 1992.

[5] Butcher, D., & Rowley, J., "The 7 Rs of Information management" Managing Information, 5(3), pp. 34-36, 1998.

[6] Shilakes, C. C., & Tylman, J., "Enterprise Information portals". New York: Merril Lynch, (November 16). [online], October 1999.

[7] Collins, D., "Data warehouses, enterprise in information portal, and the SmartMart meta directory", Information Builders Systems Journal, 12(2), pp. 53-61, 1999.

[8] Murray, G., "The Portal in the desktop". Intraspect, (May/June)[http://archives.grouping computing.com//index.cfm?fuseaction=viewarticle&contentID=166]. October 1999.

[9] Viador, "Enterprise Information portals: Realizing the vision of ourfingertips".January)[http://www.viador.com/pdfs/EIP_white_paper _1_99 .pdf]. April 2000.

[10] White, C., "Enterprise information portal requirements". Morgan Hill,CA:DatabaseaAssociatesinternational,(January)[http://www.decisio n processing.com/papers/eip2.doc]. April 2000.

[11] Wei Ma, Journal of the American Society for Information Science and Technology archive Volume 53 Issue 7, John Wiley & Sons, Inc. New York, NY, USA July 2002

[12] Dutta S.B.Wierenga and A. Dalebout, Designing Management support systems using an integrative perspective "Communication of ACM Vol. 40 No 6.1997 june.

[13] Mintzberg H., et al, The strategic Portal process, 4th. Edition upper saddle River, NJ: Prentice Hall (2002).

[14] Mora M. , Management and organizational issues(information Resources management journal special issue October/December 2002.

[15] McNurlin B.C and R.H. Information system Management , 7th. Edition upper saddle River Prentice Hall.(2005 February)

[16] Joch A.(2005,January/February)"Eye on Information" Oracle Magazine (2005,January/February)

ABOUT THE AUTHORS

Dr. (Mrs.)Philipa Omamhe Idogho, a scholar, received the Ph.D in Educational Administration from Ambrose Alli University Ekpoma, and a Master Degree in educational Management from University of Benin. With over 28 years of teaching/research experience, she is currently the Rector of a first generation Federal Polytechnic in Nigeria (Auchi Polytechnic, Auchi). She has more than 50 papers in National/International Conferences/Journals to her credit. She is a fellow, Nigeria Institute of Management, Institute of Administrative Management of Nigeria and member, Association of Business Educators of Nigeria, Association for Encouraging Qualitative Education in Nigeria, and other recognized professional organizations. Outside her academic life, she runs a Non-governmental Organisation (NGO) known as Women Enhancement Organization. It is a charitable, non-profit making, non-partisan and non-religious organization which works in three thematic areas of gender works, HIV/AIDS and literacy education for rural women and vulnerable children.

Engr. James Agajo is into a Ph.D Programme in the field of Electronic and Computer Engineering, He has a Master's Degree in Electronic and Telecommunication Engineering from Nnamdi Azikiwe University Awka Anambra State, and also possesses a Bachelor degree in Electronics and Computer Engineering from the Federal University of Technology Minna Nigeria. His interest is in intelligent system development with a high flare for Engineering and Scientific research. He has Designed and implemented the most resent computer controlled robotic arm with a working grip mechanism 2006 which was aired on a national television , he has carried out work on using blue tooth technology to communicate with microcontroller. Has also worked on thumb print technology to develop high tech security systems with many more He is presently on secondment with UNESCO TVE as a supervisor and a resource person. James is presently a member of the following association with the Nigeria Society of Engineers(NSE), International Association of Engineers(IAENG) UK, REAGON, MIRDA,MIJICTE.

Engr. Kenneth Akpado is a Ph.D Research student, he holds an M. Eng, B. Eng. in Electronics and Computer Engineering, a Member of IAENG,MNSE and COREN, he is presently a Lecturer in the Department of Electrical Electronics Engineering in Nnamdi Azikiwe University Awka Anambra State, Nigeria.

Iris Recognition Using Modified Fuzzy Hypersphere Neural Network with different Distance Measures

S. S. Chowhan
Dept. of Computer Science
COCSIT
Latur, India

U. V. Kulkarni
Dept. of Computer Science
SGGS IET
Nanded, India

G. N. Shinde
Dept. of Elec. & Computer Science
Indira Gandhi College
Nanded, India

***Abstract*—In this paper we describe Iris recognition using Modified Fuzzy Hypersphere Neural Network (MFHSNN) with its learning algorithm, which is an extension of Fuzzy Hypersphere Neural Network (FHSNN) proposed by Kulkarni et al. We have evaluated performance of MFHSNN classifier using different distance measures. It is observed that Bhattacharyya distance is superior in terms of training and recall time as compared to Euclidean and Manhattan distance measures. The feasibility of the MFHSNN has been successfully appraised on CASIA database with 756 images and found superior in terms of generalization and training time with equivalent recall time.**

Keywords- Bhattacharyya distance; Iris Segmentation; Fuzzy Hypersphere Neural Network.

I. INTRODUCTION

Iris recognition has become the dynamic theme for security applications, with an emphasis on personal identification based on biometrics. Other biometric features include face, fingerprint, palm-prints, iris, retina, gait, hand geometry etc. All these biometric features are used in security applications [1]. The human iris, the annular part between pupil and sclera, has distinctly unique features such as freckles, furrows, stripes, coronas and so on. It is visible from outside. Personal authentication based on iris can obtain high accuracy due to rich texture of iris patterns. Many researchers work has also affirmed that the iris is essentially stable over a person's life. Since the iris based personal identification systems can be more noninvasive for the users [2].

Iris boundaries can be supposed as two non-concentric circles. We must determine the inner and outer boundaries with their relevant radius and centers. Iris segmentation is to locate the legitimate part of the iris. Iris is often partially occluded by eyelashes, eyelids and shadows. In segmentation, it is desired to discriminate the iris texture from the rest of the image. An iris is normally segmented by detecting its inner areas (pupil) and outer (limbus) boundaries [3] [4]. Well-known methods such as the Integro-differential, Hough transform and active contour models have been successful techniques in detecting the boundaries. In 1993, Daugman proposed an integro-differential operator to find both the iris inner and outer borders Wildes represented the iris texture with a laplacian pyramid constructed with four different resolution levels and used the normalized correlation to determine whether the input image and the model image are from the same class [5]. O. Byeon and T. Kim decomposed an iris image into four levels using 2DHaar wavelet transform and quantized the fourth-level high frequency information to form an 87-bit code. A modified competitive learning neural network (LVQ) was used for classification [6]. J. Daugman used multiscale quadrature wavelets to extract texture phase structure information of the iris to generate a 2048-bit iris code and he compared the difference between a pair of iris representations by computing their Hamming distance [7]. Tisse. used a combination of the integro-differential operators with a Hough Transform for localization and for feature extraction the concept of instantaneous phase or emergent frequency is used. Iris code is generated by thresholding both the models of emergent frequency and the real and imaginary parts of the instantaneous phase[8]. The comparison between iris signatures is performed, producing a numeric dissimilarity value. If this value is higher than a threshold, the system generates output as a nonmatch, meaning that each iris patterns belongs to different irises [9]. In this paper, we have applied MFHSNN classifier which is a Modification of Fuzzy Hypersphere Neural Network (FHSNN) proposed by Kulkarni et al. [10]. Ruggero Donida Labati, et al. had represented the detection of the iris center and boundaries by using neural networks. The proposed algorithm starts by an initial random point in the input image, and then it processes a set of local image properties in a circular region of interest searching for the peculiar transition patterns of the iris boundaries. A trained neural network processes the parameters associated to the extracted boundaries and it estimates the offsets in the vertical and horizontal axis with respect to the estimated center [12].

II. TOPOLOGY OF MODIFIED FUZZY HYPERSPHERE NEURAL NETWORK

The MFHSNN consists of four layers as shown in Fig 1(a). The first, second, third and fourth layer is denoted as F_R, F_M, F_N, and F_o respectively. The F_R layer accepts an input pattern and consists of n processing elements, one for each dimension of the pattern. The F_M layer consists of q processing nodes that are constructed during training and each node represents hypersphere fuzzy set characterized by hypersphere membership function [11]. The processing performed by each node of F_M layer is shown in Fig 1(b). The weight between F_R and F_M layers represents centre points of the hyperspheres.

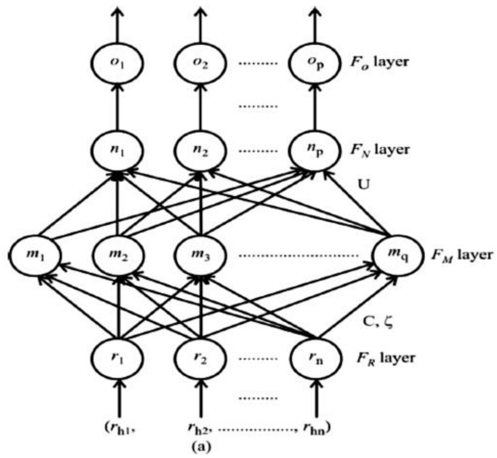


Figure 1. (a) Modified Fuzzy Hypersphere Neural Network

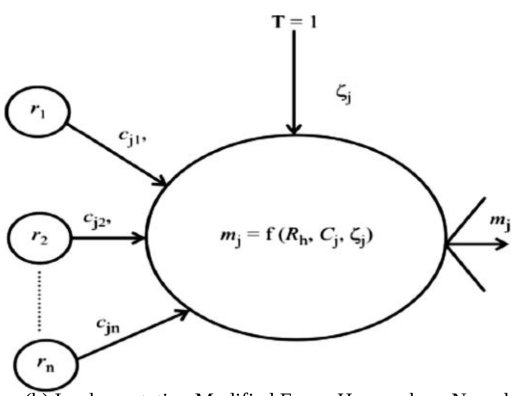


Figure1. (b) Implementation Modified Fuzzy Hypersphere Neural Network

As shown in above Fig 1(b), $C_j = (c_{j1}, c_{j2}, ...c_{jn})$ represents center point of the hypersphere m_j. In addition to this each hypersphere takes one more input denoted as threshold, T, which is set to one and the weight assigned to this link is ζ_j. ζ_j represents radius of the hypersphere, m_j which is updated during training. The center points and radii of the hyperspheres are stored in matrix C and vector ζ, respectively. The maximum size of hypersphere is bounded by a user defined value λ, where $0 \le \lambda \le l$ is called as growth parameter that is used for controlling maximum size of the hypersphere and it puts maximum limit on the radius of the hypersphere. Assuming the training set defined as $R \in \{R_k \mid h = 1, 2, ..., P\}$, where $R_h = (r_{h1}, r_{h2}, ...r_{hn}) \in I^n$ is the hth pattern, the membership function of hypersphere node m_j is

$$m_j(R_{h,} C_{j,} \zeta_j) = 1 - f(l, \zeta_j, \gamma) \quad (1)$$

Where $f(\,)$ is three-parameter ramp threshold function defined as $f(l, \zeta_j, \gamma) = 0$ if $0 \le l \le \zeta_j$, else $f(l, \zeta_j, \gamma) = l\gamma$, if $\zeta_j < l \le l$, $< l \le 1$, and argument l is defined as:

$$l = \sum_{i=1}^{n} \sqrt{\sum c_{ji} - \sum r_{hi}} \quad (2)$$

The membership function returns $m_j = 1$, if the input pattern R_k is contained by the R_k hypersphere. The parameter $\gamma, 0 \le \lambda \le l$, is a sensitivity parameter, which governs how fast the membership value decreases outside the hypersphere when the distance between R_h and C_j increases The sample plot of membership function with centre point [0.5 0.5] and radius equal to 0.3 is shown in as shown in Fig 2.

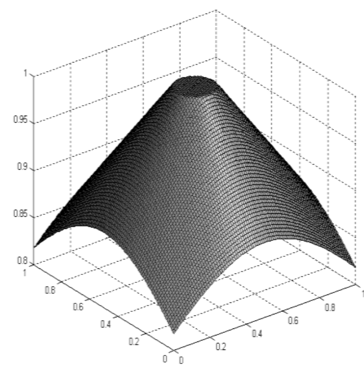


Figure 2. Plot of Modified Fuzzy Hypersphere membership function for $\gamma = 1$

Each node of F_N and F_o layer represents a class. The F_N layer gives fuzzy decision and output of kth F_N node represents the degree to which the input pattern belongs to the class n_k. The weights assigned to the connections between F_M and F_N layers are binary values that are stored in matrix U and updated during learning as, $u_{jk} = 1$ if m_j is a HS of class n_k, otherwise $u_{jk} = 0$, for $k = 1, 2, ..., p$ and $j = 1, 2, ..., q$, where m_j is the jth F_M node and n_k is the kth F_N node. Each F_N node performs the union of fuzzy values returned by HSs described as

$$n_k = \max_{j=1}^{q} m_j \, u_{jk} \;\; for \; \mathrm{k} = 1, 2 .... \mathrm{p} \quad (3)$$

Each F_o node delivers non-fuzzy output described as, $O_k = 0$, if $n_k < T$, otherwise, $O_k = 1$, if $n_k = T$, for $k = 1, 2, ..., p$ where $T = \max(n_k)$, for $k = 1, 2, ..., p$.

III. MFHSNN Learning Algorithm

The supervised MFHSNN learning algorithm for creating fuzzy hyperspheres in hyperspace consists of three steps.

A. *Creating of HSs*

Given the hth training pair (R_h, d_h) find all the hyperspheres belonging to the class d_h. These hyperspheres are arranged in ascending order according to the distances between the input pattern and the center point of the hyperspheres. After this following steps are carried sequentially for possible inclusion of input pattern R_h.

Step 1: Determine whether the pattern R_h is contained by any one of the hyperspheres. This can be verified by using fuzzy hypersphere membership function defined in equation (4). If R_h is contained by any of the hypersphere then it is included, therefore in the training process all the remaining steps are skipped and training is continued with the next training pair.

Step 2: If the pattern R_k falls outside the hypersphere, then the hypersphere is expanded to include the pattern if the expansion criterion is satisfied. For the hypersphere m_j to include R_h the following constraint must be met.

$$\left[\sum_{i=1}^{n}\sqrt{\sum c_{ji}-\sum r_{hi}}\right]\leq \lambda \qquad (4)$$

Here we have proposed a new approach for testing expansion of new hyperspheres based on Bhattacharyya distance which is sum of the absolute differences yields superior results as compared to Euclidian distance and Manhattan distance.

If the expansion criterion is met then pattern R_k is included as

$$\zeta_j^{new}=\left[\sum_{i=1}^{n}\sqrt{\sum c_{ji}-\sum r_{hi}}\right] \qquad (5)$$

Step 3: If the pattern R_k is not included by any of the above steps then new hypersphere is created for that class, which is described as

$$C_{new}=R_h \text{ and } \zeta_{new}=0 \qquad (6)$$

B. Overlap Test

The learning algorithm allows overlap of hyperspheres from the same class and eliminates the overlap between hyperspheres from different classes. Therefore, it is necessary to eliminate overlap between the hyperspheres that represent different classes. Overlap test is performed as soon as the hypersphere is expanded by step 2 or created in step 3.

Overlap test for step 2: Let the hypersphere is expanded to include the input pattern R_k and expansion has created overlap with the hypersphere, m_v which belongs to other class, which belongs to other class. Suppose, $C_u=[x_1,x_1,........x_n]$ and ζ_u represents center point and radius of the expanded hypersphere and $C_v=[x'_1,x'_2,.......x'_n]$ and ζ_v, are centre point and radius of the hypersphere of other class as depicted in Fig 3(a). Then if

$$\left[\sum_{i=1}^{n}\sqrt{\sum c_{ji}-\sum r_{hi}}\right]\leq \zeta_u+\zeta_v \qquad (7)$$

means those hyperspheres from separate classes are overlapping.

a) Overlap test for step 3

If the created hypersphere falls inside the hypersphere of other class means there is an overlap. Suppose m_p represents created hypersphere to include the input R_h and m_q represents the hypersphere of other class as shown in Fig 4b The presence of overlap in this case can be verified using the membership function defined in equation (1). If $m_q(R_h,C_p,\zeta_p)=m_q(R_h,C_q,\zeta_q)=1$ means two hyperspheres from different classes are overlapping.

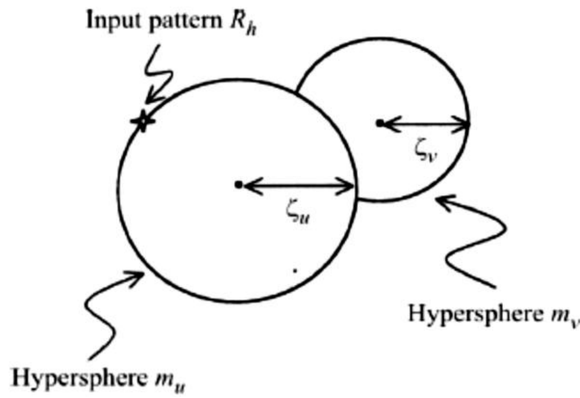


Figure 3. (a) Status of the hypersphere before removing an overlap in step 2

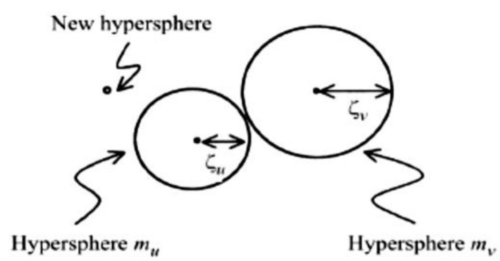


Figure 3. (b) Status of the hyperspheres after removing an overlap in step 2

b) Removing Overlap

If step 2 has created overlap of hyperspheres from separate classes then overlap is removed by restoring the radius of just expanded hypersphere. Let, m_u be the expanded hypersphere then it is contracted as

$$\zeta_u^{new}=\zeta_u^{new} \qquad (8)$$

and new hypersphere is created for the input pattern as described by equation (11). This situation is shown in Fig 3(b). If the step 3 creates overlap then it is removed by modifying of other class. Let $C_p=[x_1,x_1,........x_n]$ and ζ_p, represents centre point and radius of the $C_q=[x'_1,x'_2,x'_3,........x'_n]$ and ζ_q, are center point and radius of the

$$\zeta_j^{new} = \left[\sum_{i=1}^{n} \sqrt{\sum c_{ji} - \sum r_{hi}} \right] - \delta \qquad (9)$$

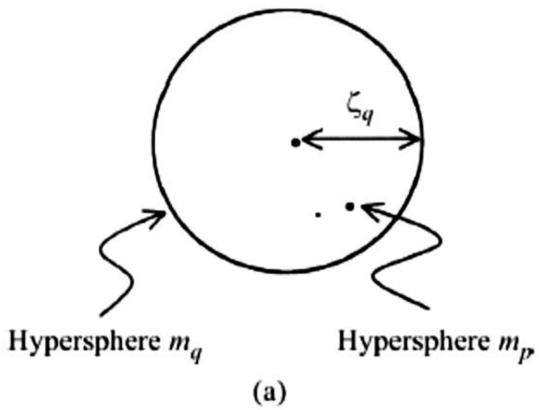


Figure 4. (a) Status of hypersphere before removing an overlap in step 3.

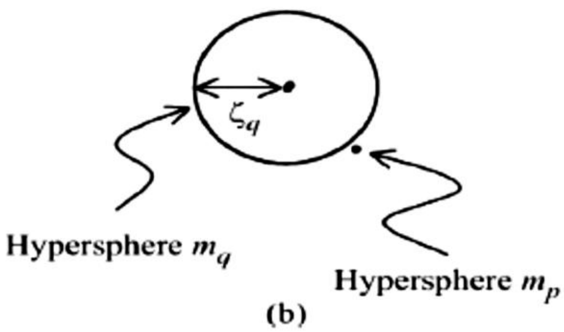


Figure 4. (b) Status of hypersphere after removing an overlap in step 3.

IV. IRIS SEGMENTATION AND FEATURE EXTRACTION

Iris Segmentation plays very important role for detecting the iris patterns, Segmentation is to locate valid part of the iris for iris biometric [3]. Finding the inner and outer boundaries (pupillary and limbic) are as shown in the Fig 5(a). Localizing it's upper and lower eyelids if they occlude, and detecting and excluding any overlaid occlusion of eyelashes and their reflection. The best known algorithm for iris segmentation is Daugman's intergro-differential operator to find boundaries of iris as defined.

$$\max(r, x_0, y_0) \left| G_\sigma * \frac{\partial}{\partial r} \int_{(r, x_0, y_0)} \frac{I(x, y)}{2\pi r} ds \right| \qquad (10)$$

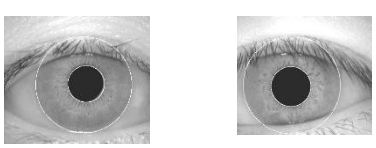


Figure 5. (a) Inner and Outer boundaries are detected with different radius

Figure 5. (b) ROI is extracted

Iris has a particularly interesting structure and provides rich texture information. Here we have implemented principal component analysis method for feature extraction which captures local underlying information from an isolated image. The result of this method yields high dimension feature vector. To improve training and recalling efficiency of the network, here we have used Singular value decomposition (SVD) method to reduce the dimensionality of the feature vector, and MFHSNN is used for Classification. SVD is a method for identifying and ordering the dimensions along which the feature exhibit the most variation [13]. The most variation are identified the best approximation of the original data points using less dimensions. This can be described as:

$$X = UTV^T \qquad (11)$$

the covariance matrix is be defined as:

$$C = \frac{1}{n} XX^T = \frac{1}{n} UT^2 U^T \qquad (12)$$

U is a $n \times m$ matrix. SVD performs repetitively order the singular values in descending order, if $n < m$, the first n columns in U corresponds to the sorted eigenvalues of C and if the first m corresponds to the sorted non-zero eigenvalues of C. The transformed data can thus be written as:

$$Y = U^T X = U^T UTV^T \qquad (13)$$

Where $U^T U$ is a simple n x m matrix which is one on the diagonal and zero. Hence Equation 13 is decomposition of equation 11.

V. EXPERIMENTAL RESULTS

CASIA Iris Image Database Version 1.0 (CASIA-IrisV1) includes 756 iris images from 108 eyes. For each eye, 7 images are captured in two sessions, where three samples are collected in the first session and four samples in second. For each iris class, we choose two samples from each session for training and remaining as testing samples. Therefore, there are 540 images for training and 216 images for testing. The timing analysis of training and recall, recognition rates in terms of number of hyperspheres, radius and recognition rate are depicted in Table 1 and Table 2.

Figure 6. Specimen Feature Vectors of 20 Iris Pattrns

TABLE I. PERFORMANCE EVALUATION OF DIFFERENT DISTANCE MEASURES ON MFHSNN

Distance Measures	Training	Recall	Hypersphere	Radius
Euclidean	0.830694	31.780012	347	0.008
Manhattan	0.867152	31.660341	364	0.022
Bhattacharya	0.750051	23.801397	377	0.0000001

TABLE II. RECOGNITION RATE OF MFHSNN ON CASIA IMAGES

Methodology		Recognition rate
Daugman	HD,SVM	99.25
Wildes	Normalized Correlation	97.43
Y.Wang	WED	99.57
Ali	HD, WED	95.20
	MFHSNN	94.00

VI. CONCLUSION AND FUTURE WORK

In this paper, we described an iris recognition algorithm using MFHSNN which has ability to learn patterns faster by creating /expanding HSs. It has been verified on CASIA database the result is as shown in Table 1 and Table II. MFHSNN can also be adapted for applications in some other pattern recognition problems. Our future work will be to improve the iris recognition rate using fuzzy neural networks.

VII. REFERENCES

[1] Li Ma, T Tan and D.Zhang and Y.Wang. *"Personal Identification Based on Iris Texture Analysis,"* IEEE Trans. Pattern Anal. Machine Intell, 25(12):1519-1533, 2003.

[2] Dey, Somnath, Samanta Debasis. "*Improved Feature Processing for Iris Biometric Authentication System,"* International Journal of Computer Science and Engineering, 4(2):127-134.

[3] Daugman. J.G, "*How Iris Recognition Works"*, Proceedings of 2002 International Conference on Image Processing, Vol. 1, 2002.

[4] Li Ma, Tieniu Tan, Yunhong Wang, and Dexin Zhang. *"Efficient Iris Recognition by Characterizing Key Local Variations",* IEEE Transactions on Image Processing, 13(6):739–750, 2004.

[5] Wildes R P. *"Iris Rcognition: An Emerging Biometric Technology"*. In Proceeding of the IEEE, Piscataway, NJ, USA, 1997, 85:1348-1363.

[6] S.Lim, K.Lee, O.Byeon and T.Kim, *"Efficient iris recognition through improvement of feature vector and classifier,"* ETRI J., 23(2): 1-70, 2001.

[7] Daugman J.G, "*Demodulation by complex-valued wavelets for stochastic pattern recognition,"* International Journal of Wavelets, Multiresolution, and Information Processing, 1(1):1–17, 2003.

[8] Tisse C.-L. and Torres L. Michel, "*Robert, Person Identification Technique Using Human Iris Recognition"*, Proceedings of the 15th International Conference on Vision Interface, 2002, pp. 294-299.

[9] Poursaberi A. and Arrabi B.N, *"Iris Recognition for Partially occluded images Methodology and Sensitive Analysis"*, Hindawi Publishing corporation journal on Advances in Signal Processing, 2007.

[10] U.V. Kulkarni and T.R. Sontakke, *"Fuzzy Hypersphere Neural Network Classifier",* Proceedings of 10th International IEEE Conference on Fuzzy Systems, University of Melbourne, Australia, December 2001.

[11] Krishna Kanth B. B. M, Kulkarni U. V. and Giridhar B. G. V., "Gene Expression Based Acute Leukemia Cancer Classification: a Neuro-Fuzzy Approach" International Journal of Biometrics and Bioinformatics, (IJBB), Volume (4): Issue (4) pp. 136–146, 2010.

[12] Ruggero Donida Labati et.al, "Neural-based Iterative Approach for Iris Detection in Iris recognition systems", Proceedings of the 2009 IEEE Symposium on Computational Intelligence in Security and Defense Applications (CISDA 2009).

[13] Madsen, R., Hansen, L., &Winther, O. "Singular value decomposition and principal component analysis," isp technical report, 2004.

AUTHORS PROFILE

Santosh S. Chowhan received the M.Sc.(CS) degree from Dr. BAM University, Aurangabad, Maharashtra, India in the year 2000. He is currently working as lecturer in the College of Computer Science and Information Technology, Latur, Maharastra. His current research interests include various aspects of Neural Networks and Fuzzy Logic, Pattern Recogntion and Biometrics

Uday V. Kulkarni received Ph.D degree in Electronics and Computer Science Engineering from S. R. T. M. University , Nanded in the year 2003. He is currently working as a professor in Dept of Computer Scince and Engineering in SGGS Institute of Engineering and Technology, Nanded, Maharashtra, India.

Ganesh N. Shinde received M. Sc. & Ph.D. degree from Dr. B.A. M. University, Aurangabad. He is currently working as Principal in Indira Gandhi College, Nanded and Maharashtra, INDIA.. He has published 27 papers in the International Journals and presented 15 papers in International Conferences. In his account one book is published, which is reference book for different courses. at Honolulu, U.S.A. He is member of Management Council & Senate of S.R.T.M. University, Nanded, INDIA. His research interest includes Filters, Image processing, Pattern recognition, Fuzzy Systems, Neural Network and Multimedia analysis and retrieval system.

Multi-Agent System Testing: A Survey

Zina Houhamdi
Software Engineering Department, Faculty of Science and IT
Al-Zaytoonah University
Amman, Jordan

***Abstract*—In recent years, agent-based systems have received considerable attention in both academics and industry. The agent-oriented paradigm can be considered a natural extension to the object-oriented (OO) paradigm. Agents differ from objects in many issues which require special modeling elements but have some similarities. Although there is a well-defined OO testing technique, agent-oriented development has neither a standard development process nor a standard testing technique. In this paper, we will give an introduction to most recent works presented in the area of testing distributed systems composed of complex autonomous entities (agents). We will provide pointers to work by large players in the field. We will explain why this kind of system must be handled differently than less complex systems.**

Keywords-Software agent; Software testing; Multi-agent system testing.

I. INTRODUCTION

As the technology evolving, the more we are driven towards abstraction and generalization. The increasing use of Internet as the spine for all interconnected services and devices makes software systems highly complex and in practice open in scale. These systems nowadays need to be adaptive, autonomous and dynamic to serve different user's community and heterogeneous platforms. These systems are developed very fast in past few decades. They are changed continuously to satisfy the business and technology modifications.

Software agents are key technologies to meet modern business needs. They offer also an efficient conceptual methodology to design such complex systems. In practice, research on software agents' development and Multi-Agent System (MAS) has become too large and used in different active area focusing mainly on architectures, protocols, frameworks, messaging infrastructure and community interactions. Thus, these systems receive more industrial attention as well.

Since these systems are increasingly taking over operations and controls in organization management, automated vehicles, and financing systems, assurances that these complex systems operate properly need to be given to their owners and their users. This calls for an investigation of appropriate software engineering frameworks, including requirements engineering, architecture, and testing techniques, to provide adequate software development processes and supporting tools.

Software agents and MAS testing is a challenging task because these systems are distributed, autonomous, and deliberative. They operate in an open world, which requires context awareness. In particular, the very particular character of software agents makes it difficult to apply existing software testing techniques to them. There are issues concerning communication and semantic interoperability, as well as coordination with peers. All these features are known to be hard not only to design and to program [3], but also to test.

There are several reasons for the increase of the difficulty degree of testing MAS:

- Increased complexity, since there are several distributed processes that run autonomously and concurrently;
- Amount of data, since systems can be made up by thousands of agents, each owning its own data;
- Irreproducibility effect, since we can't ensure that two executions of the systems will lead to the same state, even if the same input is used. As a consequence, looking for a particular error can be difficult if it is impossible to reproduce it each time [22].
- They are also non-deterministic, since it is not possible to determine a priori all interactions of an agent during its execution.
- Agents communicate primarily through message passing instead of method invocation, so existing object-oriented testing approaches are not directly applicable.
- Agents are autonomous and cooperate with other agents, so they may run correctly by themselves but incorrectly in a community or vice versa.

As a result, testing software agents and MAS asks for new testing methods dealing with their specific nature. The methods need to be effective and adequate to evaluate agent's autonomous behaviors and build confidence in them. From another perspective, while this research field is becoming more advanced, there is an emerging need for detailed guidelines during the testing process. This is considered a crucial step towards the adoption of Agent-Oriented Software Engineering (AOSE) methodology by industry.

Several AOSE methodologies have been proposed [17, 34]. While some work considered specification-based formal verification [11, 14], others borrow object-oriented testing techniques, taking advantage of a projection of agent-oriented abstractions into object-oriented constructs, UML for instance

[9, 33]. However, to the best of our knowledge, none of existing work provides a complete and *structured testing process* for guiding the testing activities. This is a big gap that we need to bridge in order for AOSE to be widely applicable.

II. SOFTWARE TESTING

Software testing is a software development phase, aimed at evaluating product quality and enhancing it by detecting errors and problems. Software testing is an activity in which a system or component is executed under specified conditions, the results are observed or recorded and compared against specifications or intended results, and an estimation is made of some aspect of the system or component. A test is a set of one or more test cases.

The principal test goal is to find faults different from errors. An error is a mistake made by the developer misunderstanding something. A fault is an error in a program. An error may lead to one or more faults. When a fault is executed then an execution error may occur. An execution error is any result or behavior that is different from what has been specified or is expected by the user. The observation of an execution error is a failure. Notice that errors may be unobservable and as a consequence may play severe disrupt with the left over computation and use of the results of this computation. The greater the period of unobserved operation, the larger is the probability of serious damage due to errors that is caused by unnoticed failures.

As showed in Figure 1, software testing consists of the dynamic verification of the program behavior on a set of suitably selected test cases. Different from static verification activities, like formal proofing or model checking, testing implies running the system under test using specified test cases [22].

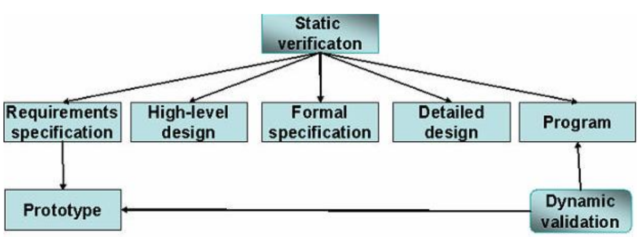


Figure 1. Kinds of Tests

There are several strategies for testing software and the goal of this survey is not to explain all of them. Nevertheless, we will describe the main strategies found in literature [22, 35]. Here they are:

- Black-box testing: also know as functional testing or specification-based testing. Testing without reference to the internal structure of the component or system.
- White-box testing: testing based on an analysis of the internal structure of the component or system. Test cases are derived from the code e.g. testing paths.
- Progressive testing: it is based on testing new code to determine whether it contains faults.
- Regressive testing: it is the process of testing a program to determine whether a change has introduced faults (regressions) in the unchanged code. It is based on reexecution of some/all of the tests developed for a specific testing activity.
- Performance testing: verify that all worst case performance and any best-case performance targets have been met.

There are several types of tests. The most frequently performed are the unit test and integration test. A unit test performs the tests required to provide the desired coverage for a given unit, typically a method, function or class. A unit test is white-box testing oriented and may be performed in parallel with regard to other units. An integration test provides testing across units or subsystems. The test cases are used to provide the needed system, as a whole, coverage. It tests subsystem connectivity. There are several strategies for implementing integration test:

- Bottom-up, which tests each unit and component at lowest level of system hierarchy, then components that call these and so on;
- Top-down, which tests top component and then all components called by this and so on;
- Big-bang, which integrates all components together;
- Sandwich, which combines bottom-up with top-down approach.

On the other hand, the goal of software testing is also to prevent defects, as it is clearly much better to prevent faults than to detect and correct them because if the bugs are prevented, there is no code to correct. This approach is used in cleanroom software development [22]. Designing tests is known as one of the best bug prevention activities. Tests design can discover and eliminate bugs at every stage in the software construction process [2]. Therefore, the idea of "test first, then code" or test-driven is quite widely discussed today. To date, several techniques have been defined and used by software developers [1].

Recently, a new testing technique called Evolutionary testing (ET) [27, 41] has been presented. The technique is inspired by the evolution theory in biology that emphasizes natural selection, inheritance, and variability. Fitter individuals have a higher chance to survive and to reproduce offspring; and special characteristics of individuals are inherited. In ET, we usually encode each test case as an individual; and in order to guide the evolution towards better test suites, a fitness measure is a heuristic approximation of the distance from achieving the testing goal (e.g., covering all statements or all branches in the program). Test cases having better fitness values have a higher chance to be selected in generating new test cases. Moreover, mutation is applied during reproduction in order to generate more different test set. The key step in ET is the transformation from testing objective to search problem, specifically fitness measure. Different testing objective gives rise to different fitness definitions. Once a fitness measure has been defined, different optimization search techniques, such as local search, genetic algorithm, particle swarm [27] can be used to generate test cases towards optimizing fitness measure (or testing objective, i.e. finding faults).

III. Software Agents and MAS Testing

A software agent is a computer program that works toward goals in a dynamic context on behalf of another entity (human or computational), perhaps for a long period of time, with discontinuous direct supervision or control, and exhibits a significant flexibility and even creativity degree in how it tries to transform goals into action tasks [18].

Software agents have (among others) the following properties:

1. Reactivity: agents are able to sense contextual changes and react appropriately;
2. Pro-activity: agents are autonomous, so they are able to select which actions to take in order to reach their goals in given situations;
3. Social ability: that is, agents are interacting entities, which cooperate, share knowledge, or compete for goal achievement

A multi-agent system (MAS) is a computational context in which individual software agents interact with each other, in a collaborative (using message passing) or competitive manner, and sometimes autonomously trying to attain their individual goals, accessing resources and services of the context, and occasionally producing results for the entities that initiated those software agents [25]. The agents interact in a concurrent, asynchronous and decentralized manner [21] hence MAS turn out to be complex systems [23]. Consequently, they are difficult to debug and test.

Due to those peculiar characteristics of agents and MAS as a whole, testing them is a challenging task that should address the following issues. (Some of them were stated in [37]):

Distributed/asynchronous: Agents operate concurrently and asynchronously. An agent might have to wait for other agents to fulfill its intended goals. An agent might work correctly when it operates alone but incorrectly when put into a community of agents or vice versa. MAS testing tools must have a global view over all distributed agents in addition to local knowledge about individual agents, in order to check whether the whole system operate accordingly to the specifications. In addition, all the issues related to testing distributed systems are applied in testing software agent and MAS as well, for example problems with controllability and observability [6].

Autonomous: Agents are autonomous. The same test inputs may result in different behaviors at different executions, since agents might modify their knowledge base between two executions, or they may learn from previous inputs, resulting in different decisions made in similar situations.

Message passing: Agents communicate through message passing. Traditional testing techniques, involving method invocation, cannot be directly applied.

Environmental and normative factors: Context and conventions (norms, rules, and laws) are important factors that govern or influence the agents' behaviors. Different contextual settings may affect the test results. Occasionally, a context gives means for agents to communicate or itself is a test input.

Scaled agents: In some particular cases, agents could be seen as scaled in that they provide no or little observable primitives to the outside world, resulting in limited access to the internal agents' state and knowledge. An example could be an open MAS that allows third-party agents to come in and access to the resources of the MAS, how do we assure that the third-party agents with limited knowledge about their intentions behave properly?

The agent oriented methodologies provide a platform for making MAS abstract, generalize, dynamic and autonomous. However, many methodologies like MASE, Prometheus, and Tropos do exist for the agent oriented framework but on contrary to it the *testing techniques* for the methodologies are not clearly supported [10].

A. Test Levels

Over the last years, the view of testing has evolved, and testing is no longer seen as a step which starts only after the implementation phase is finished. Software testing is now seen as a whole process that filters in the development and maintenance activities. Thus, each development phase and maintenance phase should have a corresponding test level. Figure 2 shows V model in which the correspondence between development process phases and test levels are highlighted [28].

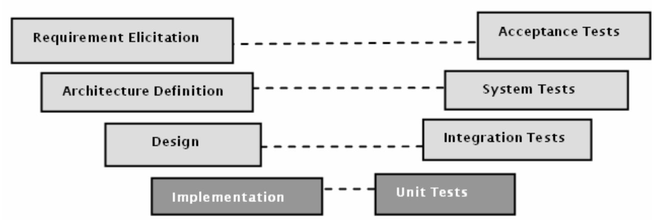


Figure 2. V model

Work in testing software agents and MAS can be classified into different testing levels: unit, agent, integration, system, and acceptance. Here we use general terminologies rather than using specific ones used in the community like group, society. Group and society, as called elsewhere, are equivalent to integration and system, respectively. The testing objectives, subjects to test, and activities of each level are described as follows:

- Unit testing tests all units that make up an agent, including blocks of code, implementation of agent units like goals, plans, knowledge base, reasoning engine, rules specification, and so on; make sure that they work as designed.
- Agent testing tests the integration of the different modules inside an agent; test agents' capabilities to fulfill their goals and to sense and effect the environment.
- Integration or Group testing tests the interaction of agents, communication protocol and semantics, interaction of agents with the environment, integration of agents with shared resources, regulations enforcement; Observe emergent properties, collective behaviors; make sure that a group of agents and environmental resources work correctly together.

- System or Society testing tests the MAS as a system running at the target operating environment; test the expected emergent and macroscopic properties of the system as a whole; test the quality properties that the intended system must reach, such as adaptation, openness, fault tolerance, performance.
- Acceptance testing tests the MAS in the customer's execution environment and verifies that it meets stakeholder goals, with the participation of stakeholders.

B. MAS Testing Problems

Defining a structured testing process for software agents and MAS: Currently, AOSE methodologies have been interesting principally on requirement analysis, design, and implementation; limited attention was given to validation and verification, as in Formal Tropos [11, 14]. A structured testing process that complements analysis and design is still absent. This problem is determinant because without detailed and systematic guidelines, the development cost may increase in terms of effort and productivity.

1. They have their own reasons for engaging in proactive behaviors that might differ from a user's concrete expectation, yet are still appropriate.
2. The same test input can give different results in different executions.
3. Agents cooperate with other agents, so they may run correctly by themselves but incorrectly in a community or vice versa.
4. Moreover, agents can be programmed to learn; so successive tests with the same test data may give different results.

As a conclusion, defining adequate and effective techniques to test software agents is, thus, a key problem in agent development.

IV. A Survey of Testing Multi-Agent Systems

There is very brief written work that describes agents software testing. The remainder of this section surveys recent and active work on testing software agents and MAS, with respect to previous categories. This classification is intended only to facilitate easily understand the research work in the field. It is also interesting to notice that this classification is incomplete in the sense that some work addresses testing in more than one level, but we put them in the level they principally focus.

A. Unit Testing

Unit testing approach calls attention to the test of the smallest building blocks of the MAS: the agents. Its essential idea is to check if each agent in isolation respects its specifications under normal and abnormal conditions. Unit testing needs to make sure that all units that are parts of an agent, like goals, plans, knowledge base, reasoning engine, rules specification, and even blocks of code work as designed. Effort has been spent on some particular elements, such as goals, plans. Nevertheless, a complete approach addressing unit testing in AOSE still opens room for research. An analogy of expected results can be those of unit testing research in the object-oriented development. At the unit level,

1. Zhang et al. [42] introduced a model based testing framework using the design models of the Prometheus agent development methodology [31]. Different from traditional software systems, units in agent systems are more complex in the way that they are triggered and executed. For instance, plans are triggered by events. The framework focuses on testing agent plans (units) and mechanisms for generating suitable test cases and for determining the order in which the units are to be tested.
2. Ekinci et al. [13] claimed that agent goals are the smallest testable units in MAS and proposed to test these units by means of test goals. Each test goal is conceptually decomposed into three sub-goals: setup (prepare the system), goal under test (perform actions related to the goal), and assertion goal (check goal satisfaction). The first and last goal prepares pre-conditions and check post-conditions while testing the goal under test, respectively. Moreover, they introduce a testing tool, called as SEAUnit that provides necessary infrastructure to support proposed approach.

B. Agent testing

At the agent level we have to test the integration of the different modules inside an agent, test agents' capabilities to achieve their goals and to sense and effect the context. There is several works in agent testing level.

1. Agile PASSI [7] proposes a framework to support tests of single agents. They develop a test suite specifically for agent verification. Test plans are prepared before the coding phase in according with specifications and the AgentFactory tool is also able of generating driver and stub agents for speeding up the test of a specific agent. Despite proposing valuable ideas concerning MAS potential levels of tests, PASSI testing approach is poorly documented and does not offer techniques to help developers in the low level design of unit test cases.
2. Lam and Barber [26] proposed a semi-automated process for comprehending software agent behaviors. The approach imitates what a human user (can be a tester) does in software comprehension: building and refining a knowledge base about the behaviors of agents, and using it to verify and explain behaviors of agents at runtime. Although the work did not deal with other problems in testing, like the generation and execution of test cases, the way it evaluates agent behaviors is interesting and relevant for testing software agents.
3. Nunez et al. [30] introduced a formal framework to specify the behavior of autonomous e-commerce agents. The desired behaviors of the agents under test are presented by means of a new formalism, called utility state machine that embodies users' preferences in its states. Two testing methodologies were proposed

to check whether an implementation of a specified agent behaves as expected (i.e., conformance testing). In their active testing approach, they used for each agent under test a test (a special agent) that takes the formal specification of the agent to facilitate it to reach a specific state. The operational trace of the agent is then compared to the specification in order to detect faults. On the other hand, the authors also proposed to use passive testing in which the agents under test were observed only, not stimulated like in active testing. Invalid traces, if any, are then identified thanks to the formal specifications of the agents.

4. Coelho et al. [8] proposed a framework for unit testing of MAS based on the use of Mock Agents. Even though they called it unit testing but their work focused on testing roles of agents at agent level according to our classification. Mock agents that simulate real agents in communicating with the agent under test were implemented manually; each corresponds to one agent role. Sharing the inspiration from JUnit [15] with Coelho et al. [8], Tiryaki et al. [40] proposed a test-driven MAS development approach that supported iterative and incremental MAS construction. A testing framework called SUnit, which was built on top of JUnit and Seagent [12], was developed to support the approach. The framework allows writing tests for agent behaviors and interactions between agents.

5. Gomez-Sanz et al. [16] introduced advances in testing and debugging made in the INGENIAS methodology [33]. The meta-model of INGENIAS has been extended with concepts for defining tests to incorporate the declaration of testing, i.e., tests and test packages. The code generation facilities are augmented to produce JUnit-based test case and suite skeletons based on these definitions with respect to debugging and it is the developer's task to modify them as needed. The work also provided facilities to access mental states of individual agents to check them at runtime. The system is integrated with ACLAnalyzer [4], a data mining facility for capturing agent communication and exploring them with different graphical representations.

6. Houhamdi [18] introduces a suite test derivation approach for Agent testing that takes goal-oriented requirements analysis artifact as the core elements for test case derivation. The proposed process has been illustrated with respect to the *Tropos* development process. It provides systematic guidance to generate test suites from agent detailed design. These test suites, on the one hand, can be used to refine goal analysis and to detect problems early in the development process. On the other hand, they are executed afterwards to test the achievement of the goals from which they were derived.

C. *Integration Testing*

Integration testing test the interaction of agents, communication protocol and semantics, interaction of agents with the context, integration of agents with shared resources, regulations enforcement; observe emergent properties; make sure that a group of agents and environmental resources work correctly together.

Only a few of methodologies define an explicit verification process by proposing a verification phase based on model checking to support automatic verification of inter-agent communications. Only some iterative methodologies propose incremental testing processes with supporting tools. At the integration level, effort has been put in agent interaction to verify dialogue semantics and workflows.

1. Agile [24] defines a testing phase based on JUnit test framework [15]. In order to use this tool, designed for OO testing, in MAS testing context, they needed to implement a sequential agent platform, used strictly during tests, which simulates asynchronous message-passing. Having to execute unit tests in an environment different from the production environment results in a set of tests that does not explore the hidden places for failures caused by the timing conditions inherent in real asynchronous applications.

2. The ACLAnalyser [4] tool runs on the JADE [39] platform. It intercepts all messages exchanged among agents and stores them in a relational database. This approach exploits clustering techniques to build agent interaction graphs that support the detection of missed communication between agents that are expected to interact, unbalanced execution configurations, overhead data exchanged between agents. This tool has been enhanced with data mining techniques to process results of the execution of large scale MAS [5].

3. Padgham et al. [32] use design artifacts (e.g., agent interaction protocols and plan specification) to provide automatic identification of the source of errors detected at run-time. A central debugging agent is added to a MAS to monitor the agent conversations. It receives a carbon copy of each message exchanged between agents, during a specific conversation. Interaction protocol specifications corresponding to the conversation are fired and then analyzed to detect automatically erroneous conditions.

4. Also at the integration level but pursuing a deontic approach, Rodrigues et al. [36] proposed to exploit social conventions, i.e. norms, rules, that prescribe permissions, obligations, and/or prohibitions of agents in an open MAS to integration test. Information available in the specifications of these conventions gives rise to a number of types of assertions, such as time to live, role, cardinality, and so on. During test execution a special agent called Report Agent will observe events and messages in order to generate analysis report afterwards.

5. Ekinci et al. [13] view integration testing of MAS rather abstract. They considered *system goals* as the source cause for integration and use them as driving

criteria. They apply the same approach for testing agent goals (unit according to their view) to test these goals. They define the concept of test goal. This concept represents the group of tests needed in order to check if the system goal is achieved correctly.

6. Nguyen et al. [29] propose using ontologies extracted from MAS under test and a set of OCL constraints, which act as a test oracle. Having as input a representation of the ontologies used, the idea is to construct an agent able to deliver messages whose content is inspired by these ontologies. The resulting behaviors are regarded as correct using the input set of OCL constraints: if the message content satisfies the constraints, the message is correct. The procedure is support by *eCAT*, a software tool.

7. Houhamdi and Athamena [19] introduced a novel approach for goal-oriented software integration testing. They propose a test suite derivation approach for integration testing that takes goal-oriented requirements analysis artifact for test case derivation. They have discussed how to derive test suites for integration test from architectural and detailed design of the system goals. These test suites can be used to observe emergent properties resulting from agent interactions and make sure that a group of agents and contextual resources work correctly together. This approach defines a structured and comprehensive integration test suite derivation process for engineering software agents by providing a systematic way of deriving test cases from goal analysis.

D. System and Acceptance Testing

System testing tests; test for quality properties, such as adaptation, openness, fault-tolerance, performance.

At the system level of testing MAS, one has to test the MAS as a system running at the target operating environment; test the expected emergent and macroscopic properties and/or the expected qualities that the intended system as whole must reach. Some initial effort has been devoting to the validation of macroscopic behaviors of MAS.

- Sudeikat and Renz [38] proposed to use the system dynamics modeling notions for the validation of MAS. These allow describing the intended, macroscopic observable behaviors that originate from structures of cyclic causalities. System simulations are then used to measure system state values in order to examine whether causalities are observable.

- Houhamdi and Athamena [20] introduced a suite test derivation approach for system testing that takes goal-oriented requirements analysis artifact as the core elements for test case derivation. The proposed process has been illustrated with respect to the *Tropos* development process. It provides systematic guidance to generate test suites from modeling artifacts produced along with the development process. They have discussed how to derive test suites for system test from late requirement and architectural design. These test suites, on the one hand, can be used to refine goal analysis and to detect problems early in the development process. On the other hand, they are executed afterwards to test the achievement of the goals from which they were derived.

Acceptance testing tests the MAS in the customer execution environment and verifies that it meets the stakeholder goals, with the participation of stakeholders. To the best of our knowledge, there is no work dealing explicitly with testing MAS at the acceptance level, currently. In fact, agent, integration, and system test harnesses can be reused in acceptance test, providing execution facilities. However, as testing objectives of acceptance test differ from those of the lower levels, evaluation metrics at this level, such as metrics for openness, fault-tolerance and adaptivity, demand for further research.

V. Conclusion

In summary, most of the existing research work on testing software agent and MAS focuses mainly on agent and integration level. Basic issues of testing software agents like message passing, distributed/asynchronous have been considered; testing frameworks have been proposed to facilitate testing process. And yet, there is still much room for further investigations, for instance:

- A complete and comprehensive testing process for software agents and MAS.
- Testing MAS at system and acceptance level: how do the developers and the end-users build confidence in autonomous agents?
- Test inputs definition and generation to deal with open and dynamic nature of software agents and MAS.
- Test oracles, how to judge an autonomous behavior? How to evaluate agents that have their own goals from human tester's subjective perspectives?
- Testing emergent properties at macroscopic system level: how to judge if an emergent property is correct? How to check the mutual relationship between macroscopic and agent behaviors?
- Deriving metrics to assess the qualities of the MAS under test, such as safety, efficiency, and openness.
- Reducing/removing side effects in test execution and monitoring because introducing new entities in the system, e.g., mock agents tester agents, and monitoring agent as in many approaches, can influence the behavior of the agents under test and the performance of the system as a whole.

References

[1] K. Beck, “Test Driven Development: By Example”, Addison-Wesley Longman Publishing Co., Boston, USA, 2005.

[2] B. Beizer, “Software Testing Techniques”, 2nd edition, Van Nostrand Reinhold Co., New York, NY, USA, 1990.

[3] F. Bergenti, M. Gleizes, and F. Zambonelli, “Methodologies and Software Engineering for Agent Systems”, The Agent-Oriented Software Engineering Handbook, Springer, Vol. 11, 2004.

[4] J. Botia, A. Lopez-Acosta, and G. Skarmeta, "ACLAnalyser: A tool for debugging multi-agent systems", Proceeding of the 16th European Conference on Artificial Intelligence, pp. 967-968, IOS Press 2004.

[5] J. Botia, J. Gomez-Sanz, and J. Pavon, "Intelligent Data Analysis or the Verification of Multi-Agent Systems Interactions", 7th International Conference of Intelligent Data Engineering and Automated Learning, Burgos, Spain, September 20-23, pp. 1207-1214, 2006.

[6] L. Cacciari, and O. Rafiq, "Controllability and observability in distributed testing", Information and Software Technology, Vol. 41, 11-12, pp. 767-780. 1999.

[7] G. Caire, M. Cossentino, A. Negri, and A. Poggi, "Multi-agent systems implementation and testing", Proceedings of the 7th European Meeting on Cybernetics and Systems Research - EMCSR2004, Vienna, Austrian Society for Cybernetic Studies, pp. 14-16, 2004.

[8] R. Coelho, U. Kulesza, A. Staa, and C. Lucena, "Unit testing in multi-agent systems using mock agents and aspects", Proceedings of the international workshop on Software engineering for large-scale multi-agent systems, ACM Press, New York, pp. 83–90, 2006.

[9] M. Cossentino, "From Requirements to Code with PASSI Methodology", In Vijayan Sugumaran (Ed.), Intelligent Information Technologies: Concepts, Methodologies, Tools, and Applications, USA, 2008.

[10] K.H. Dam, and M. Winikoff, "Comparing Agent-Oriented Methodologies", 5th International Bi-Conference Workshop, AOIS 2003 at AAMAS 2003, Melbourne, Australia, July 14, pp. 78-93, 2003.

[11] A. Dardenne, A. Lamsweerde, and S. Fickas, "Goal-directed requirements acquisition", Science of Computer Programming 20(1-2), pp.3-50, 1993.

[12] O. Dikenelli, R. Erdur, and O. Gumus, "Seagent: a platform for developing semantic web based multi agent systems", AAMAS'05 Proceedings of the fourth International Joint Conference on Autonomous agents and multi-agent systems, ACM Press, New York, pp. 1271–1272, 2005.

[13] E. Ekinci, M. Tiryaki, O. Cetin, and O. Dikenelli, "Goal-Oriented Agent Testing Revisited", Proceeding of the 9th International Workshop on Agent-Oriented Software Engineering, pp. 85-96, 2008.

[14] A. Fuxman, L. Liu, J. Mylopoulos, M. Pistore, and M. Roveri, "Specifying and analyzing early requirements in *Tropos*", Requirement Engineering, Springer Link, Vol. 9, 2, pp. 132-150, 2004.

[15] E. Gamma, and K. Beck, "JUnit: A Regression Testing Framework", http://www.junit.org, 2000.

[16] J. Gomez-Sanz, J. Botia, E. Serrano, and J. Pavon, "Testing and Debugging of MAS Interactions with INGENIAS", Agent-Oriented Software Engineering IX, Springer, Berlin, pp. 199-212, 2009.

[17] B. Henderson-Sellers, and P. Giorgini, "Agent-Oriented Methodologies", Proceedings of the 4th International Workshop on Software Engineering for Large-Scale Multi-Agent Systems - SELMAS'05, Idea Group Incorporation, 2005.

[18] Z. Houhamdi, "Test Suite Generation Process for Agent Testing", Indian Journal of Computer Science and Engineering, Vol. 2, 2, 2011.

[19] Z. Houhamdi, and B. Athamena, "Structured Integration Test suite Generation Process for Multi-Agent System", Journal of Computer Science, Vol. 7, 5, 2011.

[20] Z. Houhamdi, and B. Athamena, "Structured System Test Suite Generation Process for Multi-Agent System", International Journal on Computer Science and Engineering, Vol.3, 4, pp.1681-1688, 2011.

[21] M. Huget, and Y. Demazeau, "Evaluating multi agent systems: a record/replay approach", Intelligent Agent Technology, IAT 2004, Proceedings IEEE/WIC/ACM International Conference, pp. 536 – 539, 2004.

[22] I. Sommerville, "Software Engineering", 9th edition, Addison Wesley, 2011.

[23] N.R. Jennings, "An Agent-Based Approach for Building Complex Software Systems", Communications of the ACM, Vol. 44, 4, pp. 35-41, 2001.

[24] H. Knublauch, "Extreme programming of multi-agent systems", International Joint Conference on Autonomous Agent and Multi-Agent Systems, Bologna. ACM Press, pp. 704–711, 2002.

[25] J. Krupansky, "What is a software Agent?", Advancing the Science of Software Agent Technology, 2008. http://agtivity.com/agdef.htm.

[26] D. Lam, and K. Barber, "Debugging Agent Behavior in an Implemented Agent System", 2nd International Workshop, ProMAS, Springer, Berlin, pp. 104-125, 2005.

[27] P. McMinn, and M. Holcombe, "The state problem for evolutionary testing", Proceedings of the International Conference on Genetic and Evolutionary Computation, Springer, Berlin, pp. 2488-2498, 2003.

[28] G. Myers, "The Art of Software Testing", Wiley, 2nd Edition New Jersey, John Wiley & Sons, 2004.

[29] C. Nguyen, A. Perini, and P. Tonella, "Goal-oriented testing for MAS", Agent-Oriented Software Engineering VIII, Lecture Notes in Computer Science, Volume 4951, pp. 58-72, 2008.

[30] M. Nunez, I. Rodriguez, and F. Rubio, "Specification and testing of autonomous agents in e-commerce systems", Software Testing, Verification and Reliability, Vol. 15, 4, pp. 211-233, 2005.

[31] L. Padgham, and M. Winikoff, "Developing Intelligent Agent Systems: A Practical Guide", John Wiley and Sons, 2004.

[32] L. Padgham, M. Winikoff, and D. Poutakidis, "Adding debugging support to the Prometheus methodology", Engineering Applications of Artificial Intelligence, Vol. 18, 2, pp. 173-190, 2005.

[33] J. Pavon, J. Gomez-Sanz, and R. Fuentes-Fernandez, "The INGENIAS Methodology and Tools", In Agent Oriented Methodologies (eds. Henderson-Sellers and Giorgini), Idea group, pp. 236-276, 2005.

[34] A. Perini, "Agent-Oriented Software", Wiley Encyclopedia of Computer Science and Engineering, John Wiley and Sons, Chapter 1, pp. 1-11, 2008.

[35] R.S. Pressman, "Engenharia de Software", 6th edition, Rio de Janeiro, McGraw-Hill, 2002.

[36] L. Rodrigues, G. Carvalho, P. Barros, and C. Lucena, "Towards an integration test architecture for open MAS", 1st Workshop on Software Engineering for Agent-Oriented Systems/SBES. pp. 60-66. 2005.

[37] C. Rouff, "A test agent for testing agents and their communities", Aerospace Conference Proceedings IEEE, Vol. 5. pp. 5-2638, 2002.

[38] J. Sudeikat, and W. Renz, "A systemic approach to the validation of self-organizing dynamics within MAS", Proceeding of the 9th International Workshop on Agent-Oriented Software Engineering, pp. 237-248, 2008.

[39] TILAB. Java agent development framework. http://jade.tilab.com/.

[40] A. Tiryaki, S. Oztuna, O. Dikenelli, and R. Erdur, "Sunit: A unit testing framework for test driven development of multi-agent systems", AOSE'06 Proceedings of the 7th International Workshop on Agent-Oriented Software Engineering VII, Springer, Berlin, pp. 156-173, 2007.

[41] J. Wegener, "Stochastic Algorithms: Foundations and Applications", In Evolutionary Testing Techniques, Springer Berlin, Heidelberg, Chapter 9, pp. 82-94, 2005.

[42] Z. Zhang, J. Thangarajah, and L. Padgham, "Automated unit testing for agent systems", 2nd International Working Conference on Evaluation of Novel Approaches to Software Engineering, ENASE'07, Spain, pp. 10-18, 2007.

AUTHORS PROFILE

Dr. Zina Houhamdi received the M.Sc. and PhD. degrees in Software Engineering from Annaba University in 1996 and 2004, respectively. She is currently an Associate Professor at the department of Software Engineering, Al-Zaytoonah University of Jordan. Her research interest includes Agent Oriented Software Engineering, Software Reuse, Software Testing, Goal Oriented Methodology, Software Modeling and Analysis, Formal Methods. She has published several Research Papers in referred National/ International journals. She has referred presentations in National / International Conferences and Seminars.

Video Compression by Memetic Algorithm

Pooja Nagpal
M.Tech Student,CE deptt.
Yadavindra College of Engineering
Guru Khashi Campus,Talwandi Sabo.

Seema Baghla
Assistant Professor, CE deptt.
Yadavindra College of Engineering
Guru Khashi Campus, Talwandi Sabo

***Abstract*—Memetic Algorithm by hybridization of Standard Particle Swarm Optimization and Global Local Best Particle Swarm Optimization is proposed in this paper. This technique is used to reduce number of computations of video compression by maintaining same or better quality of video. In the proposed technique, the position equation of Standard Particle Swarm Optimization is modified and used as step size equation to find best matching block in current frame. To achieve adaptive step size, time varying inertia weight is used instead of constant inertia weight for getting true motion vector dynamically. The time varying inertia weight is based up on previous motion vectors. The step size equation is used to predict best matching macro block in the reference frame with respect to macro block in the current frame for which motion vector is found. The result of proposed technique is compared with existing block matching algorithms. The performance of Memetic Algorithm is good as compared to existing algorithms in terms number of computations and accuracy.**

Keywords- Memetic Algorithm (MA); Standard Particle Swarm Optimization (PSO); Global Local Best Particle Swarm Optimization (GLBest PSO); Video Compression; Motion Vectors; Number of Computations; Peak Signal to Noise ratio (PSNR).

I. INTRODUCTION

With the increasing popularity of technologies such as Internet streaming video and video conferencing, video compression has became an essential component of broadcast and entertainment media. Motion Estimation (ME) and compensation techniques, which can eliminate temporal redundancy between adjacent frames effectively, have been widely applied to popular video compression coding standards such as MPEG-2, MPEG-4.

Motion estimation has been popularly used in video signal processing, which is a fundamental component of video compression. In motion estimation, computational complexity varies from 70 percent to 90 percent for all video compression. The exhaustive search (ES) or full search algorithm gives the highest peak signal to noise ratio amongst any block-matching algorithm but requires more computational time [1]. To reduce the computational time of exhaustive search method, many other methods are proposed i.e. Simple and Efficient Search (SES)[1], Three Step Search (TSS)[2], New Three Step Search (NTSS)[2], Four step Search (4SS)[3], Diamond Search (DS)[4], Adaptive Road Pattern Search (ARPS)[5], Novel Cross Diamond search [6], New Cross-Diamond Search algorithm [7], Adaptive Block Matching Algorithm [8], Efficient Block Matching Motion Estimation [9], Content Adaptive Video Compression [10] and Fast motion estimation algorithm [11] . Soft computing tool such as Genetic Algorithm (GA) has also been used for fast motion estimation [12].

Traditional fast block matching algorithms are easily trapped into the local minima resulting in degradation on video quality to some extent after decoding. Since evolutionary computing techniques are suitable for achieving global optimal solution. In this paper, we propose a Memetic Algorithm to reduce number of computations of video compression by maintaining same or better quality of video. The paper is divided into five sections. The review of Particle Swarm Optimization techniques is discussed in section 2. The proposed Memetic Algorithm for video compression is discussed in section 3. Section 4 provides experimental results comparing MA with other methods for video compression. The conclusion is given in section 5.

II. PARTICLE SWARM OPTIMIZATION

The Particle Swarm Optimizer (PSO) is a population-based optimization method developed by Eberhart and Kennedy in 1995[13]. PSO is inspired by social behavior of bird flocking or fish schooling. In PSO, a particle is defined as a moving point in hyperspace. It follows the optimization process by means of local best (*Lbest*), global best (*Gbest*), particle displacement or position and particle velocity. In PSO, particle changes their positions by flying around in a multi-dimensional search space until computational limitations are exceeded. The two updating fundamental equations in a PSO are velocity and position equations. The particle velocity is expressed as Eq (1) and the particle position is expressed as Eq. (2)

$$V_{K+1}^{i} = W * V_{k}^{i} + C_1 r_1 (Lbest - S_{k}^{i}) + C_2 r_2 (Gbest - S_{k}^{i}) \quad (1)$$

$$S_{k+1}^{i} = S_{k}^{i} + V_{k+1}^{i} \quad (2)$$

Where, *V*= Particle Velocity
S= Particle Position
Lbest = Local best
Gbest = Global best
W = Inertia weight
C_1 and C_2 are acceleration constant
r_1 and r_2 are random values [0 1]
k = Current iteration

i = Particle number

In first parts, *W* plays the role of balancing the global search and local search. Second and third parts contribute to the change of the velocity. The second part of Eq. (1) is the 'cognition' part, which represents the personal thinking of the particle itself. The third part of Eq (1) is 'social part', which represents the collaboration among the particles. Without the first part of Eq. (1), all the particles will tend to move toward the same position. By adding the first part, the particle has a tendency to expand the search space, that is, they have ability to explore new area. Therefore, they acquire a global search capability by adding the first parts.

In GLBestPSO [14], inertia weight (*w*) and acceleration co-efficient (*c*) are proposed in terms of global best and local best position of the particles as given in Eq. (3) and (4). The modified velocity equation for the GLBest PSO is given in Eq. (5).

GLBest PSO is given in Eq. (5).

$$w_i = \left(1.1 - \frac{gbest_i}{pbest_i}\right) \tag{3}$$

$$c_i = \left(1 + \frac{gbest_i}{pbest_i}\right) \tag{4}$$

$$v_i(t) = w * v_i(t-1) + c_i * r(t) * (pbest_i + gbest_i - 2x_i(t)) \tag{5}$$

III. MEMETIC ALGORITHM FOR VIDEO COMPRESSION

Memetic Algorithm is developed by hybridization of Standard Particle Swarm Optimization and Global Local Best Particle Swarm Optimization. This technique is used to reduce number of computations of video compression by maintaining the same or better quality of video.

We have modified velocity and position equations of PSO to achieve step size for video compression, which is used to predict best matching macro block in the reference frame with respect to macro block in the current frame for which motion vector is found.

To get the step size, the velocity and position equations of PSO are modified as given below. The velocity equation is expressed as Eq. (6).

$$V(t) = W * C * r \tag{6}$$

Where *W* is the inertia weight, *C* is the acceleration constant, *r* is random number between 0 to 1 and *t* is generation number. To get the adaptive step size, the time varying inertia weight (*W*) is used instead of constant inertia weight similar to GLBestPSO for getting the true motion vector dynamically. The time varying inertia weight is based up on previous motion vectors as given in Eq. (7)

$$W = (1.1 - Gbest + Pbest) \tag{7}$$

$$Gbest = X + Y$$
$$Pbest = X - Y$$

Where, *X* and *Y* is the x and y coordinates of the predicted motion vector. The velocity term in Eq. (6) is added with previous motion vector to predict the next best matching block as given in Eq (8)

$$S(t+1) = S(t) + V(t) \tag{8}$$

In Memetic Algorithm, a search is made in an earlier frame of the sequence over a random area of the frame. The search is for the best matching block viz. the position that minimizes a distortion measured between the two sets of pixels comprising the blocks. The relative displacement between the two blocks is taken to be the motion vector. Usually the macro block is taken as a square of side consists of 16 pixels. The compression ration is 128:1 or 256:2. The each block size of 16 x 16 is compressed into two pixels which are nothing but motion vectors.

In Memetic Algorithm, five swarms are used to find best matching block. The initial position of block to be searched in reference frame is the predicted motion vector as expressed in Eq. (8). In Memetic Algorithm, the number of generations is taken as 2. The cost required for finding best matching block in the reference frame is ten blocks, which is less than existing methods.

The mean absolute difference (MAD) is taken as objective function or cost function in Memetic Algorithm and is expressed as in Eq. (9).

$$MAD = \frac{1}{MN}\left[\sum_{P=1}^{M}\sum_{Q=1}^{N}\left|CurrentBlock(P,Q) - ReferenceBlock(P,Q)\right|\right] \tag{9}$$

Where, *M* = Number of rows in the frame and N = Number of columns in the frame. The objective quality obtained by Memetic Algorithm has been measured by the peak signal-to-noise ratio (PSNR), which is commonly used in the objective quality comparison. The performance of the proposed method is evaluated by following Eq (10)

$$PSNR = 10\log_{10}\frac{255^2}{\frac{1}{MN}\sum_{P,Q=1}^{M,N}(OrigionalFrame(P,Q) - CompensatedFrame(P,Q))^2} \tag{10}$$

A further small improvement in the Memetic Algorithm is to check for Zero Motion Prejudgment (ZMP). If current macroblock matches with macroblock in the reference frame i.e. cost is zero then motion vector are directly stored as zero motion vector instead of gaining the motion vector through Memetic Algorithm. The zero motion prejudgment saves considerable amount of computational time.

Zero Motion Prejudgment is the procedure to find the static macro blocks which contains zero motion. In real world video sequences more than 70% of the MBs are static which do not need the remaining search. So, significant reduction of computation is possible if we predict the static macro blocks by ZMP procedure before starting motion estimation procedure and the remaining search will be faster and saves memory. We first calculate the matching error (MAD) between the macro block in the current frame and the macro block at the same location in the reference frame and then compare it to a predetermined threshold. If the matching error is smaller than predetermined threshold we consider this macro block static which do not need any further motion estimation, and return a [0, 0] as its motion vector (MV).

IV. Experimental Results And Discussions

The presented method Memetic Algorithm has been tested for two videos sequences. The performance of the proposed method is measured in terms of the *PSNR* and percentage of saving number of computations of video compression. To test the efficiency of the proposed algorithm with existing methods, the algorithms are executed in single machine. Video sequence with distance of two frames between current frame and reference frame are used to generate frame-by-frame results of the Memetic Algorithm. The performance of Memetic Algorithm is compared with other existing methods such as ES, SESTSS, TSS, NTSS and 4SS and the results are presented in Table 1 and Table 2. Figure 1 to Figure 2 shows the comparison of *PSNR* with other existing methods for two video sequences respectively. Figure 3 to Figure 4 shows the comparison of number of computations with other existing methods for two video sequences respectively. The speed of Memetic Algorithm is found to be faster than that of already published methods and *PSNR* is close to published methods as shown in Table 3 and Table 4.

The Memetic Algorithm saves number of computations up to 95.36% to 94.95 % with PSNR degradation of -0.1222 to -0.0537 as compared to ES. Similarly, Memetic Algorithm saves number of computations up to 58.95% to 35.73% with PSNR gain of +1.08 to +0.03as compared to SESTSS, TSS, NTSS and 4SS.

TABLE I. COMPARISON OF PSNR

Sequence	ES	SESTSS	TSS	NTSS	4SS	Memetic
Video 1	32.3762	31.1686	31.4379	32.2046	31.8304	32.2540
Video 2	37.2776	37.1786	37.2081	37.1939	37.1480	37.2239

TABLE II. COMPARISON OF NUMBER OF COMPUTATIONS

Sequence	ES	SESTSS	TSS	NTSS	4SS	Memetic
Video 1	207.4109	23.6349	16.2827	19.7889	18.7977	10.4642
Video 2	209.8687	23.7264	16.8776	18.7248	17.3848	9.7377

TABLE III. PSNR GAIN BY MEMETIC ALGORITHM OVER EXISTING METHODS

Sequence	ES	SESTSS	TSS	NTSS	4SS
Video 1	**-0.1222**	1.0854	0.8161	**0.0494**	0.4236
Video 2	**-0.0537**	0.0453	0.0158	**0.03**	0.0759

TABLE IV. NUMBER OF COMPUTATIONS SAVING BY MEMETIC ALGORITHM OVER EXISTING METHODS

Sequence	ES	SESTSS	TSS	NTSS	4SS
Video 1	**94.95485**	55.72564	**35.73425**	47.12086	44.33255
Video 2	**95.3601**	58.95838	**42.304**	47.99571	43.98728

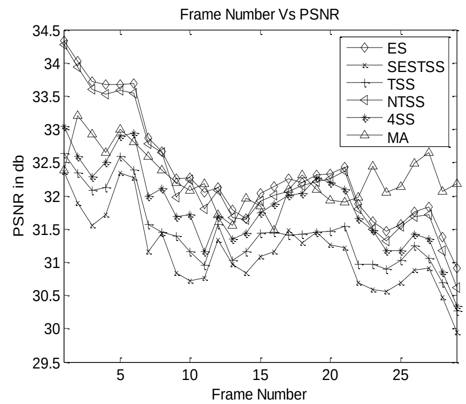


Figure1. Comparison of PSNR for Video 1

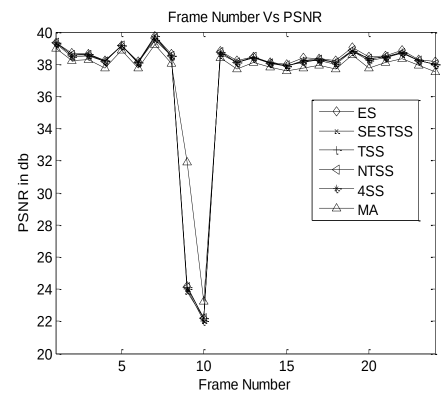


Figure2. Comparison of PSNR for Video 2

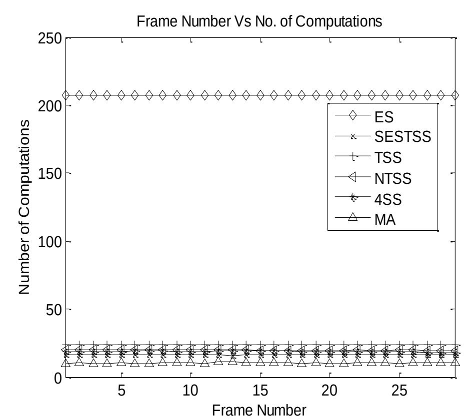


Figure3. Comparison for Number of Computations for Video 1

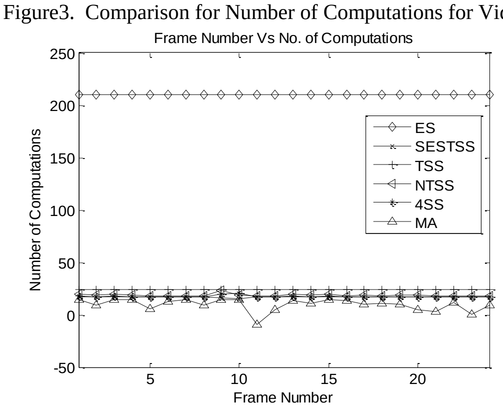


Figure 4 Comparison for Number of Computations for Video 2

V. CONCLUSION

The soft computing tool such as Memetic Algorithm is used for video compression. The performance of Memetic Algorithm is good as compared to existing algorithms except ES for video compression because it find best matching block with less computational cost by maintaining same accuracy of video. Memetic Algorithm is faster and accurate tool for video compression. The results shows promising improvement in terms of accuracy, while drastically reducing the number of computations. The step size equation of MA predicts the best matching block with less computational requirement. Memetic Algorithm uses only two-steps for video compression. An approximately more than 96% of computational time saving in the motion estimation coding is achieved as compared to ES algorithm. This saving comes with less degradation in the PSNR. This technique can easily used as block matching motion estimation algorithm or video compression algorithm.

REFERENCES

[1] Y. Wang, B. Li and T. Weise, "Estimation of Distribution and Differential Evolution Cooperation for Large Scale Economic Load Dispatch Optimization of Power Systems," *Information Sciences, Elsevier,* 2010

[2] Swagatam Das and Sudeshna Sil, "Kernel-induced fuzzy clustering of image pixels with an improved differential evolution algorithm," *Information Sciences, Elsevier,* vol. 180, pp. 1237–1256, 2010.

[3] M. Senthil Arumugam, M. V. C. Rao and Alan W.C. Tan, "A new novel and effective particle swarm optimization like algorithm with extrapolation technique," *International Journal of Applied Soft Computing, Elsevier*, vol. 9, pp. 308-320, 2009

[4] Dan Simon, "Biogeography-Based Optimization," *IEEE Trans. on Evol. Comput.* vol. 12, no.6, pp. 712-713, 2008.

[5] Jiancong Luo, Ishfog Ahmed, Yong Fang Liang and Vishwanathan Swaminathan, "Motion Estimation for Content adaptive Video Compression", *IEEE Transactions on Circuits and Systems for video Technology*, vol. 18, no.7, July 2008, pp.900-909.

[6] Chun-Man Mak, Chi keung Fong, and Wai Khen Chan, "Fast Motion Estimation For H.264/AVC in Walsh Hadmard Domain", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no.6, June 2008, pp. 735-745.

[7] M. Senthil Arumugam, M. V. C. Rao and Aarthi Chandramohan, "A New and Improved Version of Particle Swarm Optimization Algorithm with Global-Local Best Parameters," *Journal of Knowledge and Information System (KAIS), Springer*, vol. 16, no.3, pp. 324-350, 2008

[8] Srinivas Pasupuleti and Roberto Bhattiti, "The Gregarious Particle Swarm Optimizer (G-PSO)," *GECCO 2006*, Seattle, Washington, USA, July 8-12, 2006, pp. 92-98.

[9] Humaira Niasr, Tae-Sun Chol, "An Adaptive Block Motion Estimation Algorithm Based on Spatio Temporal Correlation", *Digest of Technical papers, International conference on consumer Electronics*, Jan 7-11, 2006, pp.393-394.

[10] Viet-Anh Nguyen, Yap-peng Tan, "Efficient Block Matching Motion Estimation Based on Integral Frame Attributes", *IEEE Transactions on Circuits and Systems for Video Technology*, vol.16, no. 2, March 2006, pp.375-385.

[11] Ratnaweera, S. Halgamuge and H. Watson, "Self Organizing Hierarchical Particle Swarm Optimization with time varying acceleration coefficients," *IEEE Transactions on Evolutionary Computation*, vol. 8, pp. 240-255, 2004.

[12] C.W. Lam., L.M. Po and C.H. Cheung, "A New Cross- Diamond Search Algorithm foe Fast Block Matching Motion Estimation", *2003 IEEE International Conference on Neural Networks and Signal Processing*, Nanjing, China, Dec. 2003, pp.1262-1265.

[13] Yao Nie, Kai-Khuang Ma, "Adaptive Rood Pattern Search for Fast Block-Matching Motion Estimation", *IEEE Trans. Image Processing*, vol.11, no.12,December 2002, pp. 1442-1448.

[14] Chun-Ho Cheung, Lai-Man Po, "A Novel Cross Diamond Search Algorithm for Fast Block Estimation", *IEEE Trans. Circuits and Systems*, vol.11, no.12, December 2002, pp. 1442-1448.

[15] Shan Zhu, Kai-Khuang Ma, "A New Diamond Search Algorithm for Fast Matching Motion Estimation", *IEEE Trans. Image Processing*, vol.9, no.2, February 2000, pp.287-290.

[16] Shen Li., Weipu Xu, Nanning Zheng, Hui Wang , "A Novel Fast Motion Estimation Method Based on Genetic Algorithm", *ACTA ELECTRONICA SINICA*, vol.28 no.6, June 2000, pp.114-117.

[17] Jianhua Lu., Ming L. Liou, "A Simple and Efficient Search Algorithm for Block Matching Motion Estimation", *IEEE Trans. Circuits and Systems for Video Technology*, vol.7, no.2, April 1997, pp. 429- -433.

[18] Lai-Man Po., Wing –Chung Ma, "A Novel Four- Step Search Algorithm for Fast Block Motion Estimation", *IEEE Trans. Circuits and Systems for Video Technology*, vol.6, no. 3, June 1996, pp. 313-317 .

[19] J. Kennedy and R. Eberhart, "Particle Swarm Optimization," *IEEE International Conference on Neural Networks*, Perth, Australia. vol. 4, Dec. 1995, pp. 1942-1948.

[20] Renxiang Li, Bing Zeng, and Ming L. Liou, "A New Three- Step Search Algorithm for Block Motion Estimation", *IEEE Trans. Circuits and Systems for Video Technology*, vol. 4, no. 4, August 1994, pp. 438-442.

AUTHORS PROFILE

Pooja Nagpal received diploma in Computer Engineering in 2004 from Technical Board,Chandigarh. B. Tech degree in Computer Science and Engineering in 2007 under Punjab Technical University,Jalandhar.Since 2008 she is pursuing her M.Tech from Yadavindra College of Engineering,Guru Khashi Campus,Talwandi Sabo,Bathinda(Punjab) Her interests include, Image Processing and Swarm Intelligence .She has contributed near about 8 technical papers in various national and international conferences. She is a life member of ISTE.

Seema completed M.Tech. in Computer Engineering from punjabi University Patiala in year 2007and B.Tech. in Computer Science & Engineering from Punjab Technical University, Jalandhar in year 2004 and holds NINTH merit position in University. She is presently working as Assistant Professor in Computer Engineering at Yadavindra College of Engineering, Punjabi University Guru Kashi Campus, Talwandi Sabo (Bathinda). Her research areas include Genetic Algorithms, Optimization, Image processing, motion estimation, Encryption etc.

Impact of Cloud Computing on ERP implementations in Higher Education

Ms. Shivani Goel, Dr Ravi Kiran , Dr Deepak Garg
Thapar University, Patiala

Abstract— **Penetration of Higher Education in all regions is increasing all over the globe at a very fast pace. With the increase in the number of institutions offering higher education, ERP implementations has become one of the key ingredient to achieve competitiveness in the market. Many researchers have given their inputs specifying the different nature of ERP implementation in Educational institutions then the corporate organizations. Recently Cloud computing has become a buzzword and it is having applications in many domains. Researchers have already started applying cloud computing in ERP implementations of Higher education. This paper gives an insight into the nature of cloud computing impact on ERP implementations and discusses various issues related to this. Paper comes up with guidelines regarding the use of cloud computing technology in the ERP implementations of Higher Technical Education institutions.**

Keywords- Cloud Computing; ERP; Higher Technical Education; ERP implementation.

I. Introduction

Cloud Computing has got different meaning to people working in different areas of computer science. Basically it is all about how the services are managed and delivered. Definitely it works on the principle of economies of scale at application, software and hardware level. It results

in many benefits like service provisioning, reduced costs, optimum resource utilization. Customer is now not tense about the robustness and continuity of software services in case of increasing demand of services embedded in cloud computing applications. Cloud handles elastic demand and the scalability limits go to infinity.

Good news is that good ERP systems should also have the characteristics fault tolerance, data sharing, privacy in communications and data handling, sharing data in diverse application modules.

Going a little deeper into both it will definitely strike in our mind that cloud computing technology can meet several of the advance and futuristic requirements of ERP implementations in organizations of higher technical education and can cater to increasing volume and range of services.[1]

Virtualization: It is related to autonomic computing in the context of utilizing the hardware or software resources including the processing power of a machine. There are different types of virtualizations like Hardware virtualization, Desktop virtualization, OS-level virtualization, Network virtualization and application virtualization etc. The similarity in all these is that virtualization is all about saving costs and running parallel virtual tasks from a single source increasing the utility of that source.

Service Oriented Architecture: In this kind of setting customer is the king. Customer is not aware of the heterogeneity of the different hardware architectures in the back of service providers. Customer is also not concerned with the different software applications running in different tiers of the web environment. SOA uses loose coupling concept and interfaces are defined in terms of protocols.[2]

SaaS (Software as a service) : In our context SaaS means that when the infrastructure need not to be developed in the organization but there is a service provider managing all the infrastructure at some remote location and providing the services in real time through a thin client to the users of the organization.

Ubiquitous computing: With this traditional desktop model has been left behind and users can now access the services through the device of their choice. It can be a mobile phone, ipad or any other thin client. The users are many times not even aware of this that they are using ubiquitous computing. Young generation now has grown and has left behind the desktop era.[3]

II. Evolution of ERP Implementations in Higher Education Institutions

A. *Traditional ERP implementation*

Still more than half of the institutions in the developing countries are following traditional approach due to different reasons. In these kinds of implementations different modules like admissions, fee collection, attendance, grading, Feedback system, Billing, Smart classes, Inventory, Human resource management etc are implemented in an Adhoc manner without having an overall objective of a comprehensive ERP. Many of the modules are generally developed at different times and these modules don't talk to each other. There is no concept of service architecture being used in these kinds of implementations. Cloud computing theory is out of scope for these kinds of implementations.

B. *On-Campus implementation*

Most of the implementations in the 90s stick to this category, where organisations have developed in-house infrastructure for ERP implementations and don't have any outside dependence.

C. *Implementations with Internet Host Provider*

In these kinds of implementations virtualization, Server side scripts or tools, Storage and networking component is looked after by the host who has been hired for providing the services. It leads to lower cost of hardware by the organization. It also dilutes the requirement of highly skilled professions to

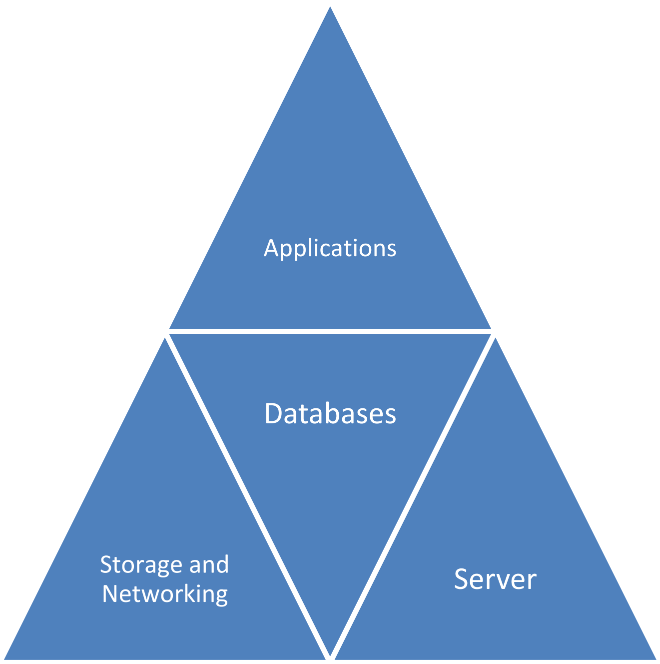


Fig 1.1 Cloud Components

maintain the system. Databases and applications are still run and managed at the campus. These implementations take some advantages of ubiquitous computing and SoA but bottlenecks still remain.[4]

D. Cloud implementations

Apart from the applications part everything is managed by the cloud server provider. Applications are anyway to be installed or made available at the user end. Applications can run in different variety of devices like TV, mobile phones, web kiosks, ipads, laptops etc. Databases, Servers, networking and storage components are still handled by the cloud provider.

III. Benefits of Cloud Computing in ERP Implementations at Higher Technical Education Institutions[5-8]

- Reduced cost: Whole of the hardware is now the responsibility of the cloud service provider. There is no requirement of any hardware at the campus, whatsoever. Even the cost of software decreases. You don't need to buy the fix time licenses or user based licenses. Now you only need to pay based on the usage of a particular service.
- Unfettered Access: All the faculty members, students, parents, management, recruiters, vendors can access the system according to the privileges granted to them. The access can be through any wired or wireless protocols using multiple devices of their choice.
- Security: Standard encryption and decryption techniques will be used and there is no need to worry about the security of the applications.
- Uptime: Almost zero downtime can be expected. However it may also depend upon the trust factor of your service provider and the feedback from the existing customers.
- Manpower: No manpower is required to be recruited at the institution.
- Futuristic Needs: Scalability, extension, peak load performance work becomes very simple. Generally this kind of support is available with every cloud provider.
- Customization: Any institution is not bound to use all the services being provided in the cloud application. The customer can choose the services based on specific requirements and the budgetary allocations.
- Group organizations: It is even more cost effective for institutions which have multiple courses and multiple campuses at different places. The access is all in real time through the web through any explorer like safari, Mozilla or IE.
- Integration : It can also integrate with the biometric, fingerprint, swipe and other machines already working in your organisations.
- There is no additional risk in case of any emergency or natural disasters because of third party involvement.The abbreviation "i.e." means "that is", and the abbreviation "e.g." means "for example".

IV. Challenges in Implementing Cloud Computing for ERP Implementation in Higher Technical Education[9-12]

- Elasticity Complexity: The demands and services in the cloud computing environment are elastic and integrated with the inherent distributed and parallel architectural characteristics. It involves highly skilled technical labour to develop and maintain such applications.
- Superstructure Emergence: Cloud is emerging as a superstructure and bypassing the boundaries of functional and vertical domains. Cloud server hosting different domain applications may become an issue because of overlapping demand and supply requirements.
- Technological Bottlenecks : Emergence of clouds also require corresponding technology upgradation in terms of new data structures for handling massive and dynamic data, new file systems, storage technologies etc.
- Serializability and consistency: Clouds are servicing the users in 24-hr format from different regions. We need to address the issue of trade-off between the consistency and serializability in sophisticated applications.

- Programming model: New Programming paradigms are need of the hour which offers something between virtualization and physical reality.
- Monitoring, Analysis and Building Trust: Due to the scale of the applications running and the criticality of the services constant monitoring of the cloud servers for different parameters of operation is required. Regular reporting after proper analysis is required, which will in turn help in building the trust of the users in cloud systems.
- Mobility and provisioning: Most of the users are now accessing the services through the devices which are fast moving. So it requires the integration of cutting edge mobile routing technologies with the cloud services technologies.

V. CONCLUSION

It has been established that cloud computing will be helpful in improving the cost, maintenance and technical efficiency of ERP implementations in technical education institutions. However, it is seen that apart from the benefits there are many issues which need to be dealt with. The issues can be overcome with proper planning. The association of right people at the time the system is evolving is very crucial.

REFERENCES

[1] Brian Hayes. 2008. Cloud computing. Commun. ACM 51, 7 (July 2008), 9-11

[2] Mache Creeger. 2009. CTO Roundtable: Cloud Computing. Queue 7, 5, Pages 1 (June 2009), 2 pages.

[3] Aaron Weiss. 2007. Computing in the clouds. netWorker 11, 4 (December 2007), 16-25

[4] Kemal A. Delic and Martin Anthony Walker. 2008. Emergence of the Academic Computing Clouds. Ubiquity 2008, August, Article 1 (August 2008)

[5] Sanjay Ghemawat, Howard Gobioff, and Shun-Tak Leung. 2003. The Google file system. In Proceedings of the nineteenth ACM symposium on Operating systems principles (SOSP '03). ACM, New York, NY, USA, 29-43.

[6] Roxana Geambasu, Cherie Cheung, Alexander Moshchuk, Steven D. Gribble, and Henry M. Levy. 2008. Organizing and sharing distributed personal web-service data. In Proceeding of the 17th international conference on World Wide Web (WWW '08). ACM, New York, NY, USA, 755-764.

[7] Michael Vrable, Stefan Savage, and Geoffrey M. Voelker. 2009. Cumulus: Filesystem backup to the cloud. Trans. Storage 5, 4, Article 14 (December 2009), 28 pages.

[8] Michael Isard and Yuan Yu. 2009. Distributed data-parallel computing using a high-level programming language. In Proceedings of the 35th SIGMOD international conference on Management of data (SIGMOD '09), Carsten Binnig and Benoit Dageville (Eds.). ACM, New York, NY, USA, 987-994.

[9] Faouzi Kamoun. 2009. Virtualizing the Datacenter Without Compromising Server Performance. Ubiquity 2009, August, pages

[10] Jeanna Matthews, Tal Garfinkel, Christofer Hoff, and Jeff Wheeler. 2009. Virtual machine contracts for datacenter and cloud computing environments. In Proceedings of the 1st workshop on Automated control for datacenters and clouds (ACDC '09). ACM, New York, NY, USA, 25-30

[11] Harold C. Lim, Shivnath Babu, Jeffrey S. Chase, and Sujay S. Parekh. 2009. Automated control in cloud computing: challenges and opportunities. In Proceedings of the 1st workshop on Automated control for datacenters and clouds (ACDC '09). ACM, New York, NY, USA, 13-18.

[12] Albert Greenberg, James Hamilton, David A. Maltz, and Parveen Patel. 2008. The cost of a cloud: research problems in data center networks. SIGCOMM Comput. Commun. Rev. 39, 1 (December 2008), 68-73.

Algorithms for Content Distribution in Networks

Dr Deepak Garg, Senior Member, IEEE
Computer Science and Engineering Department
Thapar University, Patiala

***Abstract*— In this paper an algorithm is presented which helps us to optimize the performance of content distribution servers in a network. If it is following the pay-as-you-use model then this algorithm will result in significant cost reduction. At different times the demand of different kind of content varies and based on that number of servers who are serving that demand will vary.**

Keywords- Content Distribution; Efficient Algorithms; Capacitated.

I. INTRODUCTION

The problem of content distribution in Networks is defined in the context of demand and supply paradigm. There may be downloadable software with a considerable size. At peak loads 10,000 persons may be downloading that software at the same time. However, downloads will come to an average number of 15, 00 downloads at any given point of time. Whenever the new versions of that software are launched, peak is again reached. There can be a number of such software downloads that you want to make available through your website. The demands will also be coming from different regions and different IP addresses. There will be certain pattern that can be found out by the analysis. These demand number will also keep on changing based on the time of the day in that region. The distance of the servers from that client who is requesting the software is crucial in determining the time taken to download that software. Overall performance of the site in terms of average time being taken to download one software is crucial to the image and working of the company. This is a dynamic problem, where new servers need to be made active once the active server reaches a threshold value. Similarly once there is a drop in demand from a particular region, than the server servicing that region must be relieved from service to save money. The input in this kind of problem is given in terms of a matrix containing the cost of opening of each server location and a set of locations generating the demand.[1-4]

The demands from the clients will come one by one in the form of the http request and must be handled by our algorithm. These client demands must be assigned to some server based on the location of that client and the nearest server from that position. However, it is not possible to open a new or passive demand location for a few demands. In that case these demands will be transferred to the nearest active server. When the new demands crosses a particular threshold then only new service points can be opened.

In Incremental Content Distribution Algorithm when an existing server serves number of clients less than a defined value then that server is stopped and existing demand services are transferred to another content distribution servers.

II. CONTENT DISTRIBUTION SYSTEM ALGORITHM FOR DYNAMIC AND INCREMENTAL CONTENT

Content Distribution Server problem can be characterized as:

- A universe, U, from which a set C of client input positions is selected,
- A distance metric, $d: U \times U \rightarrow R^{+}$, defined over the universe R^{+},
- An integer, $p \geq 1$, denoting the number of servers to be located, and
- An optimization function g that takes as input a set of client positions and p server positions and returns a function of their distances as measured by the metric d.

III. APPROACHES FOR HANDLING LOCATION MODELS

Formulating an appropriate model is only one step in analysing a location problem. Another much challenging task is identifying the optimal solution. Attempting a solution with well-known branch and bound optimization methods often consume unacceptable computational resources. The reason behind is that even the most basic location models are NP Hard [6]. As a result, the location analyst must have to devise other methods to identify optimal solution or at least near optimal ones. Some of the most common approaches used by location analysts are discussed below.

A. *Greedy heuristic*

A sequential approach that begins by evaluating each site individually and selecting the one facility site that yields the greatest impact on the objective. That facility site is then fixed open. The location of the next facility is then identified by enumerating all remaining possible locations and choosing the site that provides the greatest improvement in the objective. Each subsequent facility is located in an identical manner. The method stops when the required number of facilities has been sited.

B. *Improvement heuristic*

While greedy heuristics are effective at identifying a feasible solution with modest computational effort, they can't be relied upon to produce consistently good solutions. One of the earliest improvements heuristic is neighborhood search algorithm. In 1968, the most widely known improvement method was introduced [7]. The basic idea is to move a facility from the location it occupies in the current solution to an unused site. Each unused location is tried in turn and when a move produces a better objective function value, then that relocation is accepted and we have a new (improved) solution. The search procedure is repeated on the new solution and stops when no better solution can be found via this method. A variable neighborhood search algorithm was presented for solving the p-median problem [8]. The algorithm performs an intensive local search on the current solution until it settles in local optima. It then repeats the process by randomly selecting a solution from a neighborhood at a distance k from the current best solution. The process continues, increments k, until some predefined maximum value of k is attained.[9-11]

C. Lagrangian relaxation

When using any heuristic we are trading on savings in solution time against the quality of the solution while the heuristic often find good solutions to a variety o location problems, it is difficult to evaluate the trade off since we have no way of knowing how far from optimality those solutions are. Without having the optimal value of the objective function available for comparison, we can sometimes approximate the difference between a heuristic's solution and the optimal solution by finding bounds. One of the primary attraction of the technique known as Lagrangian relaxation is that it provides upper and lower bound on the value of objective function [12]. This is done by eliminating i.e relaxing one or more of the constraint of the original model and adding these constraints multiplied by an associated Lagrange multiplier to the objective function. The role of these multipliers is to derive the Lagrangian problem towards a solution that satisfies the relaxed constraints. The primary challenge in applying such technique is in selecting which constraint to relax. Ideally the relaxed problem ought to be solvable by inspection or by a simple sorting the objective function coefficients.

Algorithm

Input : Cost of starting a new server and a queue of client requests {cr1,cr2,…,crn}

A set S is maintained which consists of currently active servers to process that demand. The set will look like {φ , s1,s2,…sn}. Sets of Client requests which are being handled by individual servers will also be maintained.

Step 1: When a new client request for a download to start, either it can be assigned to the existing active server or a new server should be started.

Step 2. Let si is the server nearest to the client request and number of client requests assigned to si< threshold value then the download request is assigned to that server

Step 3: If si is the server nearest to the client request and number of client requests assigned to si=threshold value then the new server needs to be assigned. New server will be started in the cloud of regions being represented by the incoming request for download.

Step 4: Another thread will keep a check on the number of requests that are being services by a server because some of them will be completed at any given time.

Step 5: Once the number of requests are below a threshold then it will check for the possibility of transferring the existing download request of one server to another server keeping in mind that the total load does no increases beyond the limits after such transfer.

The Total cost will be calculated as

(Number of active servers) * cost of starting a server + Sum of all client requests for download taken over all distances of existing servers

The evaluation of this algorithm tells that it will give results close to the optimum values that can be reached.

IV. Results and Discussion

By the nature of the problem it can be seen that essentially it an NP-complete problem in which the number of options can be exponential. Because a particular request canbe assigned to any of the active set of n servers giving rise to n different options, similarly if there are m different requests that will come to exponential possibilities. By the above algorithm, the solution will be close to the actual solution and limited to 1.5 times the value of the exact solution.

V. Conclusion

Content management algorithms initially were only dealing with static data. Now improved algorithms can also handle dynamic nature of the data, locations, demands and resources. The resultant saving is significant in terms of time and cost. There is also a possibility of applying the concepts of linear programming to further improve the approximation ratio.

References

[1] N. Alon, B. Awerbuch, Y. Azar, N. Buchbinder, and J. Naor. A General Approach to Online Network Optimization Problems. ACM Transactions on Algorithms, 2(4):640 to 660, 2006.

[2] V. Arya, N. Garg, R. Khandekar, A. Meyerson, K. Munagala, and V. Pandit. Local Search Heuristics for k-Median and Facility Location Problems. SIAM J. on Computing, 33(3):544 to 562, 2004.

[3] A. Borodin and R. El-Yaniv. Online Computation and Competitive Analysis. Cambridge University Press, 1998.

[4] M. Charikar and R. Panigrahy. Clustering to Minimize the Sum of Cluster Diameters. In Proc. of the 33rd ACM Symposium on Theory of Computing (STOC '01), pages 1 to 10, 2001.

[5] M. Chrobak and C. Kenyon-Mathieu. Online Algorithms Column 10: Competitiveness via Doubling. SIGACT News, 37(4):115 to 126, 2006.

[6] J. Fakcharoenphol, S. Rao, and K. Talwar. Approximating Metric Spaces by Tree Metrics. Encyclopedia of Algorithms, pages 51 to 53, 2008.

[7] D. Fotakis and C. Tzamos. Winner-Imposing Strategyproof Mechanisms for Multiple Fa-cility Location Games. In Proc. of the 6th Workshop on Internet and Network Economics (WINE '10), volume 6484 of LNCS, pages 234 to 245, 2010.

[8] G. Frahling and C. Sohler. Coresets in Dynamic Geometric Data Streams. In Proc. of the 37th ACM Symposium on Theory of Computing (STOC '05), pages 209 to 217, 2005.

[9] P. Lu, X. Sun, Y.Wang, and Z.A. Zhu. Asymptotically Optimal Strategy-Proof Mechanisms for Two-Facility Games. In Proc. of the 11th ACM Conference on Electronic Commerce (EC '10), pages 315 to 324, 2010.

[10] R.R. Mettu and C.G. Plaxton. The Online Median Problem. SIAM J. on Computing,32(3):816 to 832, 2003.

[11] A. Meyerson. The Parking Permit Problem. In Proc. of the 46th IEEE Symposium on Foundations of Computer Science (FOCS '05), pages 274-284, 2005.

[12] R.L. Francis (Editors) P.B. Mirchandani. Discrete Location Theory. Willey, 1990.

[13] B. Krishnamurthy, C. Wills, Y.Zhang On the use and performance of content distribution networks IMW'01 poceedings of the 1st ACM SIGCOMM Workshop on Internet Measurement,2001

[14] G.Pallis, A.Vakali Insight and perspectives for content delievery networks , Communications of the ACM, 2006

AUTHORS PROFILE

Dr Deepak Garg, Senior Member of ACM and is Chair, ACM North India SIGACT. He is Senior Member IEEE and Secretary IEEE Delhi Section Computer Society. He has more than 50 publications to his credit. His research interests are Algorithm design and complexity issues. He is having 15 years of research experience

Generation of Attributes for Bangla Words for Universal Networking Language(UNL)

Muhammad Firoz Mridha
Dept. of Computer Science
Stamford University Bangladesh
Dhaka, Bangladesh

Kamruddin Md. Nur
Dept. of Computer Science
Stamford University Bangladesh
Dhaka, Bangladesh

Manoj Banik
Dept. of Computer Science and Engineering
Ahsanullah University of Science and Technology,
Dhaka, Bangladesh

Mohammad Nurul Huda
Dept. of Computer Science and Engineering
United International University
Dhaka, Bangladesh

***Abstract*— The usage of native language through Internet is highly demanding now a day due to rapidly increase of Internet based application in daily needs. It is important to read all information in Bangla from the internet. Universal Networking Language (UNL) addressed this issue in most of languages. It helps to overcome the language barrier among people of different nations to solve problems emerging from current globalization trends and geopolitical interdependence. In this paper we propose a work that aims to contribute with morphological analysis of those Bangla words from which we obtain roots and Primary suffixes and developing of grammatical attributes for roots and Primary suffixes that can be used to prepare Bangla word dictionary and Enconversion/Deconversion rules for Natural Language Processing(NLP).**

Keywords- Morphology; Bangla Roots; Primary Suffix; Grammatical Attributes; Universal Networking Language.

I. Introduction

In the last few years, machine translation techniques have been applied to web environments. The growing amount of available multilingual information on the Internet and the Internet users has led to a justifiable interest on this area. Hundreds of millions of people of almost all levels of education and attitudes, of different jobs all over the world use the Internet different purposes [1]. English is the main language of the Internet. Understandably, not all people know English. Urgent need, therefore, arose to develop interlingua translation programs. The main goal of the UNL system is to provide Internet users access to multilingual websites using a common representation. This will allow users to visualize websites in their native languages. The UNL system has a growing relevance since the usage of the WWW is generalized across cultural and linguistic barriers. Many languages [10] such as Arabic, French, Russian, Spanish, Italian, English, Chinese or Brazilian Portuguese have already been included in the UNL platform. Lexical knowledge representation is a critical issue in natural language processing systems. Recently, the development of large-scale lexica with specific formats capable of being used by several different kinds of applications has been given special focus; in particular to multilingual systems. Our aim is to introduce Bangla into this system. In order to implement this project with the lowest time and human effort costs, we will reuse linguistic resources already available as much as possible.

In this paper we present the UNL system for Bangla. The major components of our research works touches upon i) development of grammatical attributes for Bangla root and Krit Prottoy to construct *Bangla Word Dictionary* and use of morphological analysis ii) UNL Expression of the Bangla attributes and *iii)* Selecting scheme of attributes. In section 2 we describe the UNL system. In sections 3, 4 and 5 we present our main works that include all the above three components.

II. Universal Networking Language

UNL is an artificial language that allows the processing of information across linguistic barriers [10]. This artificial language has been developed to convey linguistic expressions of natural languages for machine translation purposes. Such information is expressed in an unambiguous way through a semantic network with hyper-nodes. Nodes (that represent concepts) and arcs (that represent relations between concepts) compose the network. UNL contains three main elements:

• Universal Words: Nodes that represent word meaning.

• Relation Labels: Tags that represent the relationship between Universal Words. Tags are the arcs of UNL hypergraph.

• Attribute Labels: Additional information about the universal words.

These elements are combined in order to establish a hierarchical Knowledge Base (UNLKB) [10] that defines unambiguously the semantics of UWs. The UNL Development Set provides tools that enable the semi-automatic conversion of natural language into UNL and vice-versa. Two of such tools are the EnConverter and the DeConverter. The main role of EnConverter [11] is to translate natural language sentences into UNL expressions. This tool implements a language independent parser that provides a framework for morphological, syntactic and semantic analysis synchronously. This allows morphological and syntactical ambiguities resolution. The DeConverter [3, 12], on the other hand, is a language independent generator that converts UNL expressions to natural language sentences.

A. *Universal Words*

Universal Words are words that constitute the vocabulary of UNL. A UW is not only a unit of the UNL syntactically and semantically for expressing a concept, but also a basic element for constructing a UNL expression of a sentence or a compound concept. Such a UW is represented as a node in a hypergraph. There are two classes of UWs from the viewpoint in the composition:

• labels defined to express unit concepts and called "UWs" (Universal Words)
• a compound structure of a set of binary relations grouped together and called "Compound UWs".

B. *Relational Labels*

The relation [1] between UWs is binary that have different labels according to the different roles they play. A relation label is represented as strings of three characters or less. There are many factors to be considered in choosing an inventory of relations. The following is an example of relation defined according to the above principles.

Relation: There are 46 types of relations in UNL. For example, *agt* (agent), *agt* defines a thing that initiates an action, agt(do, thing), agt(action, thing), obj(thing with attributes) etc.

C. *Attributes*

The attributes represent the grammatical properties of the words. Attributes of UWs are used to describe subjectivity of sentences. They show what is said from the speaker's point of view: how the speaker views what is said. This includes phenomena technically [4, 5] called speech, acts, propositional attitudes, truth values, etc. Conceptual relations and UWs are used to describe objectivity of sentences. Attributes of UWs enrich this description with more information about how the speaker views these state of affairs and his attitudes toward them.

III. MORPHOLOGY OF BANGLA WORDS

Morphology is the field of linguistics that studies the structure of words. It focuses on patterns of word formation within and across languages, and attempts to formulate rules that model the knowledge of the speakers of those languages. Thus morphological analysis is found to be centered on analysis and generation of word forms. It deals with the internal structure of words and how words can be formed. Morphology plays an important[2, 8] role in applications such as spell checking, electronic dictionary interfacing and information retrieving systems, where it is important that words that are only morphological variants of each other are identified and treated similarly. In natural language processing (NLP) and machine translation (MT) systems we need to identify words in texts in order to determine their syntactic and semantic properties [7]. Morphological study comes here to help with rules for analyzing the structure and formation of the words. A Bangla morpheme, besides the root word, is supposed to be represented in the Bangla-UNL dictionary using the following UNL format [10].

[HW] "UW" (ATTRIBUTE 1, ATTRIBUTE 2 …) <FLG, FRE, PRI>
HW← Head Word (Bangla Word)
UW← Universal Word
ATTRIBUTE← Attribute of the HW
FLG← Language Flag
FRE← Frequency of Head Word
PRI← Priority of Head Word

The attributes describe the nature of the head word classifying it as a grammatical, semantic or morphological feature. So, we will be especially concerned about representation of morphemes using various attributes.

A. *Prefixes (উপসর্গ)*

Prefixes are the words that are used before words to express various meanings of the same words. There are around fifty (50) prefixes used in Bangla sentences. In Shangskrit Bangla we use twenty (20) prefixes[2] say , প্র(প্রকর্ষ), পরা(বৈপরীত্য), অপ(বৈপরীত্য) etc. , in Bangla we use thirteen prefixes (13) prefixes[2] such as বে(বৈপরীত্য), গর(বৈপরীত্য), অন(অভাব) etc., five(5) foreign prefixes[5] such as গর (না),বদ (খারাপ) etc., four English prefixes[14] such as সাব(অধীন অর্থে), হেড(প্রধান), ফুল(পুরা), হাফ(অর্ধ) etc. These prefixes are used before words to make thousands of meaningful Bangla Words.

In our work, we will make separately Word Dictionary entries for all of these prefixes and words, so that they can combinely make meaningful words by applying rules. For example, if we consider prefix "প্রতি"[9] (means like/similar/every/opposite/against etc.) we can make "প্রতিদিন", "প্রতিশব্দ" etc. Now we can make the word "প্রতি" for dictionary entry. But the word "প্রতি" has two or more meanings so that we have to represent two or more dictionary entries for the word as follows.

[প্রতি]{} "every (icl>thing)" (ABSTRACT THING) <B,0,0>
[প্রতি]{} "opposite (icl>thing)" (ABSTRACT THING) <B,0,0>

Now if we want to represent the concepts of the words say প্রতিদিন, প্রতিশব্দ etc., we need not represent the whole words. We have to represent only the words "দিন", "শব্দ" in the dictionary entry as per the following format.

[দিন]{}"day (icl>period>time)" (N, ABSTRACT THING, LIGHT) <B, 0, 0>
[শব্দ] {} "sound (icl>occure>thing)" (N, ABSTRACT THING) <B, 0, 0>

B. Noun Morphology

Bangla Nouns have very strong and structural inflectional morphology base on case. Case of noun may be nominative (“ছেলে”, boy), accusative (ছেলে-কে”), to the boy) and genetive (“ছেলে-র”, of the boy) and so on. Gender and number are also important for identifying proper categories of nouns. Number may be singular (“ছেলে”, boy or “ছেলেটি”, the boy, “বই”, book, “বইটি”, the book) plural (“ছেলেরা”, boys “ছেলেগুলি”, the boys “বইগুলো”, the books etc.). So, from the word “ছেলে” we get “ছেলের”, “ছেলেকে”, “ছেলেটি”, “ছেলেগুলি” etc. and from the word “বই” we get “বইটি”, “বইগুলি” etc. Some dictionary entries may look like. [ছেলে]{} “boy (icl>person)” (N, HN, C, ANI)<B,0,0>

Here, “boy (icl>person)” is the UW for “ছেলে” but “র”, “কে” etc. have no UWs. Therefore, they should be represented in the dictionary only with grammatical attributes as follows.

[র] {} “” (3P, SUF, N)<B,0,0>

[কে] {} “” (3P, SUF, N, HUMN, SG) <B,0,0>

[রা] {} “” (3P, PL, SUF, N, HUMN) <B,0,0>

[টি] {} “” (N, SG, SUF,3P) <B,0,0>

[গুলি] {} “” (N, PL, SUF,3P) (<B,0,0>

C. Adjective Morphology

As Adjective we can consider Bangla words “সাহস”, “সুন্দর” and “**ভাল**” meaning “bravery”, “beautiful” and “good” in English respectively. From the first word we get সাহসী (সাহস+ই), সাহসের (সাহস+এর). And from the second and third words we get সুন্দরী, ভালোর, ভালোটা etc. We have to have the dictionary entries for সাহস, সুন্দর, **ভালো,** ই, এর, র, টা to make the meaningful words সাহসী, সাহসের, সুন্দরী, ভালোটা etc. by combining the morphemes with the root words using analysis rules.

D. Pronoun Morphology

Here we can consider the word root “তাহা”(he/she). From this we get তাহা-রা, তাহা-কে, তাহা-দের, তাহা-দের-কে, তাহা-দিগকে etc. So, we have to consider these morphemes রা, কে, দের, দিগকে for dictionary entries to form words with “তাহা” as above.

E. Verb Morphology

Diversity of verb morphology in Bangla is very significant. We can select the head words as the Longest Common Lexical Unit (LCLU) of all the possible transformations of the word [8].

We can give the example of the Bangla word “পড়” (means read). The corresponding UW in basic form is “read”. The dictionary entry is: [পড়া] { } “read (icl>do)”, where ‘পড়v &’ is the *head word* and (icl>do) is from the knowledge base. Some possible transformations of “পড়" in the Bangla to UNL dictionary are given as follows [9, 10]:

If we consider ‘পড়’ (means read) as a root, we can represent this root in the dictionary as

[পড়]{} “read (icl>do)” (V, @present) <B,0,0>

Some transformations based on the persons and tenses are.

For first person:
[পড়] { } “read (icl>do)” (ROOT, BANJANT)<B, 0, 0>
[ই] { } “read (icl>do)” (ROOT, BANJANT, PRESENT INDEF)<B, 0, 0>
[ইতেছি] { } “read (icl>do)” (ROOT, BANJANT, PRESENT CONT)<B,0, 0>

For second person:
[পড়] { } “read (icl>do)” (ROOT, BANJANT)<B, 0, 0>
[ইতেছ] { } “read (icl>do)” (ROOT, BANJANT, PRESENT CONT)<B, 0, 0>
[ইবে] { } “read (icl>do)” (ROOT, BANJANT, FUTURE INDEF)<B, 0, 0>

For third person:
[পড়] { } “read (icl>do)” (ROOT, BANJANT)<B, 0, 0>
[ইয়াছে] { } “read (icl>do)” (ROOT, BANJANT, PRESENT PERF)<B, 0, 0>
[ইতেছে] { } “read (icl>do)” (ROOT, BANJANT, PRESENT CONT)<B, 0, 0>

For resolving the ambiguities of the words গিয়েছি, গিয়েছিলাম, গিয়েছেন, গিয়েছিলেন, যাইতে থাকবে, etc. we have to define them as full words for dictionary entries. For instance [গিয়েছিলাম] {} “go (icl>move>do)”(V, PAST, INDEF, 1P). Using the same procedure we can make dictionary entries for different transformations of other roots such as কর (do), লিখ (write),দে (give) etc.

IV. Morphology of Bangla Root Words

Bangla Language contains a lot of verbs. The core part of those verbs is called roots. In another way if we split the verbs we get two parts *Roots* and *Suffixes*. From verbs if we remove suffixes we get roots. For example ‘K‡i’ (do) is a verb. Its two parts are: Ki+G; here ‘Ki’ is a root and ‘G’ is a suffix. Some other Bangla roots verbs are Pj& , co&, ai& , Mo&, Nl&, bvP&, Kuv`& etc.

A. Bangla Primary Suffixes (K…ৎ cÖZ¨q)

We know that the core of the verb is called root and if number of suffixes are added to roots then they form verbs. When sound or sounds [8] are added with roots and form nouns or adjectives then the root words are called root verbs and the sound or sounds are added with root verbs are called Primary Suffixes. For example Pj& (Root verb)+Ab& (Primary Suffix)=Pjb (Noun) and Pj& (Root verb)+Aš— (Primary Suffix)= Pjš— (Adjective). Some others primary suffixes are Ab, Abv, Awb, AK, Av etc.

B. Vowel ended and Consonant ended root

Verb roots that are ended with vowel are called SORANTO root. For example, চা, হ, পা, নী, ধু, খা, ছা etc. And verb roots that are ended with consonant are called BANJANTO root. For example, কর্,,চল্ ,পড়্,নাচ্,চিন্.etc.

C. Morphological Analysis of Bangla verbs

Morphological analysis is applied to identify the actual meaning of the word by identifying suffix or morpheme of that word. Every word is derived from a root word. A root word may have the different transformations. This happens because of different morphemes which are added with it as suffixes. So, the meaning of the word varies for its different transformations. For example, if we consider 'Ki&'(do) as a root word then after adding ÕB' we get the word 'Kwi'[6, 8] which means a work done by some one(first person , present tense). Similarly after adding 'Av', we get the word 'Kiv'. Here, this word represents noun of the root word 'Ki&'. Therefore, by morphological analysis we get the grammatical attributes of the main word. Derivational morphology is simple and a word rarely uses the derivational rule in more than two or three steps. The first step forms nouns or adjectives from verb roots. The next steps form new nouns and adjectives [5]. We have examined derivational morphology for UNL Bangla dictionary too.

V. Methods of Finding Grammatical Attributes

Representing Universal Words (UWs) for each of the Bangla Head Word we need to develop grammatical attributes that describe how the words behave in a sentence. Grammatical Attributes (GA) have to be developed by the rules (Enconversion and Deconversion) and dictionary developers. They play very important rules for writing Enconversion and Deconversion rules because a rule uses GA in morphological and syntactic analysis, to connect or analyze one morpheme with another to build a meaningful (complete) word and to examine or define the position of a word in a sentence.

When we analyze the HWs for representing them in the word dictionary as UWs, we find all the possible specifications of the HWs as attributes named grammatical attributes, so that they can be used in the dictionary for making rules (EnCo and DeCo). For example, if we consider "cvwLÓ meaning bird as a head word, then we can use attributes N (as it is noun), ANI (as bird is an animal), SG for singular number and CONCRETE (as it a concrete thing which is touchable).

So, this word can be represented in the dictionary as follows:

[cvwL]{}"bird(icl>animal>animate thing)"(N,ANI,SG,CONCRETE)

Head Word | Universal Word | Grammatical Attributes

Same way we can represent the words avb (paddy), bvP& (dance) etc. as follows:

[avb] {} "paddy(icl>plant>thing)"(N, PLANT, CONCRETE)
[bvP&]{} "dance(icl>do)"(ROOT, BANJANT)

Where, N for noun, CONCRETE for any concrete thing which is touchable, ANI for animal and SG for singular number.

As we are the initiators of developing rules and word dictionary for Bangla we are proposing some grammatical attributes and their descriptions in table 1.

A. Primary suffixes with consonant ended roots (BANJANTO):

Some primary suffixes that are added with roots (which are ended with consonant) to form nouns and adjectives are given in table 2.

Table 2 Consonant ended roots with primary suffixes

Banjanto root + Primary suffix	Noun and Adjective
কর্+অন্ত	করন্ত,
পড়্+অন্ত	পড়ন্ত,
চল্+অন্ত	চলন্ত
চল্+তি	চলতি
পড়্+তি	পড়তি
কাট্+তি	কাটতি
ঝুল্+তি	ঝুলতি

The dictionary entries for the roots and their suffixes are as follows:

[নাচ]{} "do (icl>do)"(ROOT, BANJANT)\
[পড়্]{} "read (icl>do)"(ROOT, BANJANT)
[চল্]{} "go (icl>do)"(ROOT, BANJANT)
[কাট্]{} "cut (icl>do)"(ROOT, BANJANT)
[পাকড়্]{} "attack (icl>do)"(ROOT, BANJANT)
[উড়্]{} "fly (icl>do)"(ROOT, BANJANT)
[অন]{} "" (PROT, KPROT, BANJANT, MVNOUN)
[আ]{} "" (PROT, KPROT, BANJANT, MNOUN)
[অন্ত]{} "" (PROT, KPROT, BANJANT, MADJ)

B. Primary suffix with vowel ended roots (SORANTO):

Some SORANTO is added only with *primary suffix* 'Av' to form nouns are shown in table 3.

Table 3 Vowel ended roots with primary suffix

Soranto root + Primary suffix	Noun
চা+আ	চা ওয়া
পা+আ	পাওয়া
দে+আ	দে ওয়া
খা+আ	খা ওয়া

So, the dictionary entries would be:

[চা]{} "want(icl>do)" (ROOT, SORANT) <B,0,0>

[পা]{} “get(icl>obtain)” (ROOT, SORANT) <B,0,0>
[দে]{} “give(icl>do)” (ROOT, SORANT) <B,0,0>
[খা]{} “eat(icl>do)” (ROOT, SORANT) <B,0,0>
[আ]{}“” (PROT, KPROT, SORANT, MNOUN) <B,0,0>

C. There are some such primary suffixes which have two forms. One form is added with vowel ended roots and other forms are added with consonant ended roots. Those are given in table 4.

Table 4 Primary suffix with SORANTO roots and BANJANTO roots

Banjanto root + Primary suffix	Noun
চা+আ	চা ওয়া
পা+আ	পাওয়া
দে+আ	দে ওয়া
চল্+আ	চলা
পড়্+আ	পড়া
কাট্+আ	কাটা

The UNL form:

[চা]{} “want(icl>do)” (ROOT, SORANT) <B,0,0>
[পা]{} “get(icl>obtain)” (ROOT, SORANT) <B,0,0>
[দে]{} “give(icl>do)” (ROOT, SORANT) <B,0,0>
[আ]{}“”(PROT,KPROT,SORANT,MNOUN)<B,0,0>
[পড়]{} “read (icl>do)”(ROOT, BANJANT) <B,0,0>
[চল্]{} “go (icl>do)”(ROOT, BANJANT) <B,0,0>
[কাট্]{} “cut (icl>do)”(ROOT, BANJANT) <B,0,0>
[আ]{}“”(PROT,KPROT,BANJANT,MNOUN)<B,0,0>

All the Roots are added with “আ” Krit Prottoy and form a meaningful word. For example,

কর্+আ=করা
হার্+আ=হারা

But Some BANJANTA Roots (দুল্,খুল্) etc do not form a meaningful Bangla word to add with “আ” Prottoy. For example,

দুল্+আ=দুলা
খুল+আ=খুলা.

Rather দোল in place of দুল্ and খোল in place of খুল is added with “আ” Krit Prottoy and form meaningful word. For example, দোল+আ=দোলা, খোল +আ=খোলা.

Grammatical Attributes	Descriptions	Examples (Here we can use Bangla Words)
ROOT	Only for verb root	co& (Read), Ki&(do) etc.
SORANT	Verb roots that are ended with vowel	চা (want),যা(go) etc.
BANJANT	Verb roots that are ended with consonant	পড়(read), ai& (catch) etc.
1P	First person	vwg (I), Avgiv (we) etc.
2P	Second Person	Zzwg (you) etc.
3P	Third Person	‡m cyi“l (He), ‡m gwnjv (She) etc.
PROT	For all suffixes	Av, Ab, AvB etc.
KPROT	For the suffixes that are used after roots to create Nouns, Adjectives etc.	BK, Ab etc.
KBIVOKTI	For the suffixes that are used after roots to create only verbs.	B, B‡ZwQ, †e etc.
ADJ	For any adjective	fvj (good), my›`i (beautiful) etc.
N	For any noun	Kjg (pen), Avg(mango) etc.
MNOUN	suffixes that are added with roots to make nouns.	Av etc.
MADJ	suffixes that are added with roots to make adjectives	Aš— etc.
MVERB	suffixes that are added with roots to form verbs	B, B‡ZwQ, †e etc.
UROOT	For consonant উ	দুল্, খুল etc.
PRESENT INDEFINITE	suffixes that create present indefinite form of the sentence	B etc.
HPRON	For human	আমি, সে etc.
NUM	For number	৫,৭,৯ etc.
NANI	For not amimate	বই, কলম etc.
CONCRETE	For solidthing	জমি, ঘর etc.
HF	For human food	ভাত, রুটি etc.
FEMALE	For female person	সে(মহিলা) etc.
BIV	For Bivokti	ইতেছে,ইয়াছে
V	For Verb	করে,পড়ি etc
3PG	3rd Person General	করিতেছে, করিত etc
IMPS	Imperative Sentence	করুক,করুন etc
3PR	3rd Person Respected	করিতেছেন, করিবেন etc
2PR	2nd Person Respected	করিতেছেন, করিবেন etc
2PG	2nd Person General	করিতেছ, করিতেছিল etc
2PN	2nd Person Neglected	কর, করিস etc
PAH	Past Habituate	করিত, করিতেন etc
1P	1st Person	করি, করিতেছি etc

Table 1 some proposed grammatical attributes

D. To solve this problem we divide BANJANTA Roots into two categories. One is General BANJANTA that is attributed as BANJNT and another is attributed with URoots.

Table 5 Examples of general BANJANTO and URoots

Banjanto root + Primary suffix	Noun
কর্+আ	করা
পড়্+আ	পড়া
হার্+আ	হারা
দুল্+আ	দুলা
খুল+আ	খুলা
দোল+আ	দোলা
খোল+আ	খোলা

The UNL form:

[কর্]{} “do (icl>do)”(ROOT, BANJANT) <B,0,0>
[হার্]{} “lost (icl>do)”(ROOT, BANJANT) <B,0,0>
[আ]{} (PROT, KPROT, SORANT, MNOUN) <B,0,0>
[দোল]{} “swing (icl>do)”(ROOT, BANJANT) <B,0,0>
[খোল]{} “open (icl>do)”(ROOT, BANJANT) <B,0,0>
[দুল্] {} “swing(icl>do)”(ROOT, BANJNT ,URoot) <B,0,0>
[খুল] {} “open(icl>do)”(ROOT, BANJNT ,URoot) <B,0,0>

We make two entries in Bangla word dictionary for URoots. For example, দুল্ and দোল, খুল and খোল. In Bangla there are some primary suffixes which added with roots and form a new different words for example, বচ + w^3 = মু w^3 and মুচ+ w^3= মু w^3. These words are added to word dictionary in special category.

The dictionaries would be:

[উw^3]{} “speech (icl>do)”(ROOT, BANJANT, SP) <B,0,0>
[মুw^3] {} “free (icl>do)”(ROOT, BANJANT, SP) <B,0,0>

The suffixes আন্ত, তি, আ etc. will be in the dictionary only with grammatical attributes. They will be added with the roots to form verbs, nouns or adjectives using rules. In the above examples we have classified suffixes in the basis of adding either with SORANTO (vowel ended) or BANJANTO (consonant ended) to give them proper attributes so that they can be used to make appropriate rules enconversion and deconversion.

VI. CONCLUSION AND FUTURE WORK

A system capable of understanding natural language sentences is of potentially unlimited uses in the field of natural language processing. In this paper we have generated grammatical attributes of Bangla roots and Krit Prottoy for developing Bangla Word Dictionary for Universal Networking Language (UNL).

We have presented some method to select grammatical attribute using morphological analysis Bangla words that can be used to make dictionary for converting the Bangla sentences to UNL documents and vice versa. We have done limited work so far for Bangla words.

Our future plan is to build a Bangla language server that will contain a complete Bangla Word Dictionary.

REFERENCES

[1] S. Abdel-Rahim, A.A. Libdeh, F. Sawalha, M. K. Odeh, “Universal Networking Language(UNL) a Means to Bridge the Digital Divide”, Computer Technology Training and Indistrial Studies Center, Royal Scientific Sciety, March 2002.

[2] M. M. Asaduzzaman, M. M. Ali, “Morphological Analysis of Bangla Words for Automatic Machine Translation”, International Conference on Computer, and Information Technology (ICCIT), Dhaka, 2003, pp.271-276

[3] Serrasset Gilles, Boitel Christian, (1999) UNL-French econversion as Transfer & Generation from an Interlingua with Possible Quality nhancement through Offline Human Interaction. *Machine Translation Summit-VII*, Singapore.

[4] M. E. H. Choudhury, M. N.Y. Ali, M.Z.H. Sarkar, R. Ahsan, “Bridging Bangla to Universal Networking Language- A Human Language Neutral Meta- Language”, International Conference on Computer and Information Technology (ICCIT), Dhaka, 2005,pp.104- 109

[5] M.E.H. Choudhury, M.N.Y. Ali, “Framework for synthesis of Universal Networking Language”, East West University Journal, Vol. 1, No. 2, 2008, pp. 28-43

[6] M.N.Y. Ali, J.K. Das, S.M. Abdullah Al Mamun, M. E.H. Choudhury, ”Specific Features of a Converter of Web Documents from Bengali to Universal Networking Language”, International Conference on Computer and Communication Engineering 2008(ICCCE’08), Kuala Lumpur, Malaysia.pp. 726-731

[7] S. Dashgupta, N. Khan, D.S.H. Pavel, A.I. Sarkar, M. Khan, ”Morphological Analysis of Inflecting Compound words in Bangla”, International Conference on Computer, and Communication Engineering (ICCIT), Dhaka, 2005, pp. 110-117

[8] M.N.Y. Ali, J.K. Das, S.M. Abdullah Al Mamun, A. M. Nurannabi,”Morpholoical Analysis of Bangla worfs for Universal Networking Language”,icdim,08.

[9] Bangla Academy (2004), Bengali-English Dictionary, Dhaka.

[10] H. Uchida, M. Zhu, “The Universal Networking Language (UNL) Specification Version 3.0”, Technical Report, United Nations University, Tokyo, 1998

[11] Enconverter Specifications, version 3.3, UNL Center/ UNDL Foundation, Tokyo, Japan 2002.

[12] Deconverter Specifications, version 2.7, UNL Center/ UNDL Foundation, Tokyo, JapanA.

[13] D. M. Shahidullah, “Bangala Vyakaran”, Maola Brothers Prokashoni, Dhaka, August 2003,pp.110-130

[14] D. C. Shniti Kumar,”Vasha-Prokash Bangla Vyakaran”, Rupa and Company Prokashoni, Calcutta, July 1999, pp.170-175

AUTHORS PROFILE

Muhammad Firoz Mridha is now a full time faculty in Stamford University Bangladesh. He obtained M.Sc. in Computer Science and Engineering from United International University (UIU) in 2010 and B.Sc. in Computer Science Engineering 2004 from Khulna University of Engineering and Technology(KUET). His research interest includes Natural Language Processing, Artificial Intelligence and Software.

Kamruddin Md. Nur attended Victoria University of Wellington, New Zealand (VUW) for his bachelor in Computer Science (2004) and United International University (UIU), Bangladesh for his masters in Computer Science (2010). He is a permanent faculty member of Computer Science department at Stamford University Bangladesh and his research interests include – Software Engineering Approaches, Software Visualization, Web Engineering, Ubiquitous and Mobile Computing, Usability Engineering, Artificial Intelligence etc.

Manoj Banik was born in Habiganj, Bangladesh in 1972. He completed his B.Sc. Engineering in Computer Science and Engineering from BUET, Dhaka, Bangladesh and M.Sc. Engineering in Computer Science and Engineering from UIU, Dhaka, Bangladesh. Now, he is working as an Assistant Professor in the Department of Computer Science and Engineering of Ahsanullah University of Science and Technology, Dhaka,

Bangladesh. His research interests include Speech Recognition, Phonetics, Universal Networking Language and Algorithms. He is a member of Bangladesh Engineers Institute (IEB).

Mohammad Nurul Huda was born in Lakshmipur, Bangladesh in 1973. He received his B. Sc. and M. Sc. in Computer Science and Engineering degrees from Bangladesh University of Engineering & Technology (BUET), Dhaka in 1997 and 2004, respectively. He also completed his Ph. D from the Department of Electronics and Information Engineering, Toyohashi University of Technology, Aichi, Japan. Now he is working as an Associate Professor in United International University, Dhaka, Bangladesh. His research fields include Phonetics, Automatic Speech Recognition, Neural Networks, Artificial Intelligence and Algorithms. He is a member of International Speech Communication Association (ISCA).

A Modified Method for Order Reduction of Large Scale Discrete Systems

Dr. G.Saraswathi
Dept. of EEE, GIT, GITAM University
Visakhapatanam, India

***Abstract*— In this paper, an effective procedure to determine the reduced order model of higher order linear time invariant discrete systems is discussed. A new procedure has been proposed for evaluating Time moments of the original high order system. Numerator and denominator polynomials of reduced order model are obtained by considering first few redefined time moments of the original high order system. The proposed method has been verified using numerical examples.**

Keywords- order reduction; eigen values; large scale discrete systems; modeling.

I. INTRODUCTION

The rapid advancements in science and technology led to extreme research in large scale systems. As a result the overall mathematical complexity increases. The computational procedure becomes difficult with increase in dimension. Therefore high-order models are difficult to use for simulation, analysis or controller synthesis. So it is not only desirable but necessary to obtain satisfactory reduced order representation of such higher order models. Here the objective of the model reduction of high order complex systems is to obtain a Reduced Order Model(ROM) that retains and reflects the important characteristics of the original system as closely as possible.

Several methods are available in the literature for large-scale system modeling. Most of the methods based on the original continued fraction expansion technique [1, 2] fail to retain the stability of the original systems in the reduced order models. To overcome this major drawback, alternative methods have been suggested but the common limitation of such extension is that in some cases, they may generate models of order even higher than that of the original system [1-3].

Modal-Padé methods [4] use the concept of the dominant poles and matching the few initial time moments of the original systems. The major disadvantage of such methods is the difficulty in deciding the dominant poles of the original system, which should be retained in the reduced order models. Retaining the poles closest to the imaginary axis need not be always the best choice. Sinha et al, [5] used the clustering technique but the serious limitation of this method is the number and position of zeros of the original system sometimes decides the minimum order of the reduced order model. The major disadvantage of such methods is in deciding the clusters of poles hence cannot generate unique models.

Some methods based on Eigen Spectrum which is the cluster of poles of high order system considered to derive the approximant. Pal *et al* proposed [6] pole-clustering using Inverse Distance Criterion and time-moment matching. Vishwakarma and R.Prasad [7,8] modified the pole clustering by an iterative method. The difficulty with these methods is in selecting poles for the clusters. Mukherjee [9], suggested a method based on Eigen Spectrum Analysis. The Eigen Spectrum consists of all poles of the high order system. The poles of the reduced model are evenly spaced between the first and last poles. Parmar *et al* [10] proposed a mixed method using Eigen Spectrum Analysis with Factor division algorithm to determine the numerator of the reduced model with known denominator. Parmar *et al* [11] proposed another mixed method using Eigen Spectrum Analysis equation and Particle Swarm method to find the reduced model. The methods based on eigen spectrum analysis cannot be applied to high order systems having complex poles. Saraswathi *et al* [17] proposed a method retaining some of the properties of original system based on eigenspectrum. The method can be applied for systems having both real and complex poles unlike the other existing methods [9-11].

Many methods are available for order reduction of high order continuous systems but very few are extended for discrete systems. The proposed method is extended to discrete systems using Tustin approximation [12] to maintain the static error constants identical in both discrete and continuous transfer functions. In the proposed reduced order method the Time moments and Markov parameters are redefined as function of Residues and poles for strictly proper rational transfer functions having real and complex poles. Poles of the Reduced Order Model (ROM) are selected by considering the highest contribution of each pole in redefined Time Moments (RTMs) and lowest contribution in Redefined Markov Parameters (RMPs).

II. PROPOSED METHOD

Let the original high order transfer function of a linear time invariant discrete system of n^{th} order be

$$G(z) = \frac{N(z)}{D(z)} = \frac{c_m z^m + c_{m-1} z^{m-1} + \cdots \ldots + c_1 z + c_0}{z^n + d_{n-1} z^{n-1} + d_{n-2} z^{n-2} + \cdots \ldots + d_1 z + d_0}; \; m < n \quad \ldots(1)$$

Let the original high order transfer function of a linear time invariant discrete system of n^{th} order in continuous form using Tustin approximation be

$$G(s)=\frac{N(s)}{D(s)}=\frac{b_m s^m+b_{m-1}s^{m-1}+\cdots\ldots+b_1 s+b_0}{s^n+a_{n-1}s^{n-1}+a_{n-2}s^{n-2}+\cdots\ldots+a_1 s+a_0};\ m<n \quad ...(2)$$

where $D(s)=\prod_{i=1}^{n}(s+\lambda_i)$; n is the number of poles and $N(s)=\prod_{i=1}^{m}(s+\delta_i)$; m is the number of zeros, $\lambda_i=(\alpha_i \pm j\beta_i)$ and $\delta_i=(\varepsilon_i \pm j\gamma_i)$ of high order system $G(s)$. The poles and zeros may be real and/or complex. If they are complex, they occur in conjugate pairs.

The reduced order model of k^{th} order using proposed new algorithm in continuous form is defined as

$$R_k(s)=\frac{N_k(s)}{D_k(s)}=\frac{p_r s^r+p_{r-1}s^{r-1}+\cdots\ldots+p_1 s+p_0}{s^k+q_{k-1}s^{k-1}+q_{k-2}s^{k-2}+\cdots\ldots+q_1 s+q_0};\ k<n \quad \ldots (3)$$

where $D_k(s)=\prod_{i=1}^{k}(s+\acute{\lambda}_i)$; k is the number of poles and $N_k(s)=\prod_{i=1}^{m}(s+\acute{\delta}_i)$; r ≤ m, r is the number of zeros, $\acute{\lambda}_i=(\acute{\alpha}_i \pm j\acute{\beta}_i)$ and $\acute{\delta}_i=(\acute{\varepsilon}_i \pm j\acute{\gamma}_i)$ of reduced order model $R_k(s)$. In the reduced model poles and zeros may be real and/or complex. If they are complex, they occur in conjugate pairs as mentioned for original system.

Using Tustin approximation the reduced order linear time invariant discrete model of k^{th} order is

$$R_k(z)=\frac{N_k(z)}{D_k(z)}=\frac{u_r s^r+u_{r-1}s^{r-1}+\cdots\ldots+u_1 s+u_0}{s^k+v_{k-1}s^{k-1}+v_{k-2}s^{k-2}+\cdots\ldots+v_1 s+v_0};\ k<n \quad ...(4)$$

We know that the power series expansion of $G(s)$ about $s = 0$ is

$$G(s)=C_0+C_1 s+C_2 s^2+\cdots=\sum_{i=0}^{\infty}C_i s^i \quad \ldots(5)$$

Where $C_i=\frac{(-1)^i}{i!}RTM_i$; i = 0,1,2,3, …..

The expansion of $G(s)$ about $s = \infty$ is

$$G(s)=\mu_0 s^{-(n-m)}+\mu_1 s^{-(n-m+1)}+\mu_2 s^{-(n-m+2)}+\cdots$$
$$=\sum_{i=0}^{\infty}\mu_i s^{-(n-m+i)} \quad \ldots(6)$$

Where $\mu_i=RMP_i$; I = 0,1, 2, 3, …..

i) Considering the original high order system $G(s)$ with distinct poles

Define the expressions for redefined time moments (RTMs) as

$$RTM_i=\sum_{j=1}^{n}x_{ij}\ ; i=0,1,2,\ldots. \quad \ldots(7)$$

where $x_{ij}=i!\frac{P_j}{\lambda_j^{(i+1)}}; i=0,1,2,\ldots.$

Define Redefined Markov Parameters (RMPs) as

$$RMP_i=\sum_{j=1}^{n}y_{ij} \quad \ldots(8)$$

where $y_{ij}=P_j\lambda_j^i; i=0,1,2,\ldots.$ and P_j are residues.

ii) If $G(s)$ is having 'r' repeated poles

$$G(s)=\frac{b_m s^m+b_{m-1}s^{m-1}+\cdots\ldots+b_1 s+b_0}{(s+\lambda_1)^r(s+\lambda_{r+1})\ldots\ldots(s+\lambda_n)}\ ;\ m<n$$

Define the expressions for redefined time moments (RTMs) as

$$RTM_i=i!\left[\frac{P_1}{\lambda_1^{(i+1)}}+[ET_{ix}]+\sum_{k=r+1}^{n}\frac{P_k}{\lambda_k^{(i+1)}}\right]; i=0,1,2,. \quad \ldots(9)$$

where $ET_{ix}=\left[\sum_{x=1}^{r-1}\frac{P_{(x+1)}}{\lambda_1^{(i+x+1)}}\left\{\prod_{j=1}^{x}\frac{(-1)^j(i+j)}{x!}\right\}\right]$ and P_j are residues.

Define Redefined Markov Parameters (RMPs) as

$$RMP_i=\left[P_1\lambda_1^i+[EM_{ix}]+\sum_{k=r+1}^{n}P_k\lambda_k^i\right];\ i=0,1,2,.. \quad \ldots(10)$$

where $EM_{ix}=\left[\sum_{x=1}^{r-1}P_{(x+1)}\lambda_1^{(i-x)}\left\{\prod_{j=1}^{x}\frac{(i-j+1)}{x!}\right\}\right]$; E_{ix}= 0 if $i \le x$ and P_j are residues.

The denominator polynomial $D_k(s)$ of the k^{th} order reduced model is obtained by selecting poles with the highest contribution in RTMs and lowest contribution in RMPs according to their contribution weight age as shown in Table I.

Table I : Contributions of individual poles

Parameters	λ_1	λ_2	λ_3	…	λ_j	…	λ_n	Sum
RTMs	x_{i1}	x_{i2}	x_{i3}	…	x_{ij}	…	x_{in}	RTM_i
RMPs	y_{i1}	y_{i2}	y_{i3}	…	y_{ij}	…	y_{in}	RMP_i

Where x_{ij} is the contribution of pole λ_j in RTM_i and y_{ij} is the contribution of pole λ_j in RMP_i. The numerator polynomial, $N_k(s)$ of the k^{th} order reduced model is obtained by retaining the first few initial RTMs and RMPs of the original system as follows:

$$N_k(s)=\sum_{i=0}^{r_1}p_i s^i+\sum_{i=1}^{r_2}p_{(r_1+i)}s^{(r_1+i)};\ r=r_1+r_2,\ r\le m$$
and $r_1 \ge 1$. …(11)

where $p_i=\sum_{i=0}^{j}q_i C_{(j-i)}$; j = 0, 1, 2, 3,…, r_1 ; r_1 is number of RTMs …(12)

and $p_{(r-j)}=\sum_{i=0}^{j}q_{(r-i)}\mu_{(j-i)}$; j = 0, 1, 2, 3,…, r_2 ; r_2 is number of RMPs …(13)

r_2, the number of RMPs is zero if r_1 is considered. If $n-m<1$, naturally r_2 = 0.

III. EXAMPLE

Considering a sixth order discrete system described by the transfer function given as

$$G_1(z)=\frac{0.3277z^6+0.9195z^5+1.038z^4+0.5962z^3+0.1618z^2+0.006986z-0.005308}{z^6+1.129z^5+0.2889z^4-0.08251z^3-0.04444z^2-0.00476z}$$

Using the Tustin transformation $G_1(z)$ is transformed to $G_1(s)$ with sampling time T =1.

$$G_1(s)=\frac{s^5+15.6s^4+124.2s^3+510.3s^2+1166s+959.3}{s^6+21s^5+175s^4+735s^3+1624s^2+1764s+720}$$

The poles of $G_1(s)$ are $-\lambda_1$ = -1, $-\lambda_2$ = -2, $-\lambda_3$ = -3, $-\lambda_4$ = -4, $-\lambda_5$ = -5 and $-\lambda_6$ = -6. The contributions of individual poles are derived from equ.(7) for RTMs and from equ.(8) (non zero terms) for RMPs. These contributions are tabulated in Tables II and III. Poles having highest contribution in RTMs and lowest contribution in RMPs according to their contribution weight age are - λ_{z_1} = -1 and - λ_{z_3} = -4.

Table II Contribution of poles in RTMs and RMPs of HOS

RTMs/RMPs	λ_1	λ_2	λ_3
RTM_0=1.3324	1.6167	2.2396	-7.7444
RTM_1=1.6448	1.6167	1.1198	-2.5815
RMP_0=1.0000	1.6167	4.4792	-23.233
RMP_1= -5.4	-1.6167	-8.9583	69.700

Table III Contribution of poles in RTMs and RMPs of HOS

λ_4	λ_5	λ_6
10.8146	-8.4433	2.8493
2.7036	-1.6887	0.4749
43.2583	-42.2167	17.0958
-173.0333	211.0833	-102.575

For the second order model with two poles in ESZ, Denominator polynomial is

$$D_2(s) = (s + 1)(s + 4)$$
$$= s^2 + 5s + 4$$

Numerator of the ROM is obtained using equation (11) matching first two RTMs of the original system from Table II.

$$N_2(s) = (1.3324 \text{ X } 4) + \{(1.3324 \text{ X } 5) + (-1.6448 \text{ X } 4)\}s$$
$$= 5.329 + 0.8244s$$

The transfer function of the second order reduced model is

IV. $\boldsymbol{R_{1-1}(s)} = \frac{\mathbf{0.8244s+5.329}}{\mathbf{s^2+5s+4}}$

The conversion of continuous transfer function to discrete $R_{1-1}(s)$ to $R_{1-1}(z)$ is done using the Tustin transformation with Sampling time 1. The second order reduced discrete model is

$$R_{1-1}(z) = \frac{0.3052z^2 + 0.5922z + 0.2869}{z^2 - 0.1111}$$

The second order model by Farsi et al [13] is

$$R_{1-1}^F(z) = \frac{0.589z - 0.4495}{z^2 - 1.428z + 0.5329}$$

The step response of the proposed second order discrete model $R_{2\text{-}1}(z)$ and second order discrete model of Farsi *et al* [13] $R^F_{2\text{-}1}(z)$ are compared with original discrete system $G_1(z)$ in Fig.1. The step response by proposed method is following the original very closely when compared Farsi *et al*[13].

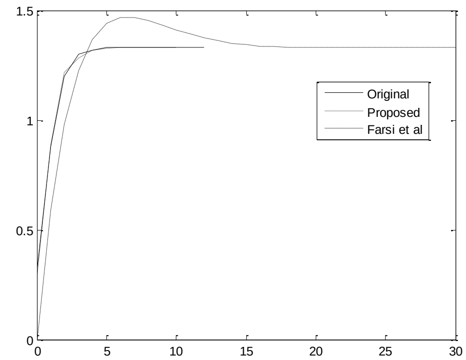


Fig.1 Comparison of step responses of $G_1(z)$, $R^F_{2\text{-}1}(z)$ and $R_{2\text{-}1}(z)$

V. CONCLUSIONS

Conclusions and Future Research: In this paper, an effective procedure to determine the reduced order model of higher order linear time invariant discrete systems is presented. Numerator and denominator polynomials of reduced order model are obtained by redefining the time moments of the original high order system. The stability of the original system is preserved in the reduced order model as the poles are taken from the original system. The method produces a good approximation when compared with other methods. The method is applied for real, complex and repeated poles of continuous [14] and discrete systems and the work is in progress to make it generalize for interval systems.

REFERENCES

[1] Sinha, N.K. and Pille,W., "A new method for order reduction of dynamic systems", International Journal of Control, Vo1.14, No.l, pp.111-118, 1971.

[2] Sinha, N.K. and Berzani, G. T., "Optimum approximation of high-order systems by low order models", International Journal of Control, Vol.14, No.5, pp.951-959, 1971.

[3] Davison, E. J., "A method for simplifying linear dynamic: systems", IEEE Trans. Automat. Control, vol. AC-11, no. 1, pp.93-101, 1966.

[4] Marshall, S. A., "An approximate method for reducing the order of a linear system", International Journal of Control, vol. 10, pp.642-643, 1966.

[5] Mitra, D., "The reduction of complexity of linear, time-invariant systems", Proc. 4th IFAC, Technical series 67, (Warsaw), pp.19-33, 1969.

[6] J.Pal, A.K.Sinha and N.K.Sinha, "Reduced order modeling using pole clustering and time moments matching", Journal of the Institution of engineers (India), Pt.EL, Vol pp.1-6,1995.

[7] C.B.Vishwakarma, R. Prasad, "Clustering methods for reducing order of linear systems using Padé Approximation", IETE Journal of Research, Vol.54, Issue 5, pp.326-330, 2008.

[8] C.B.Vishwakarma, R. Prasad, "MIMO system reduction using modified pole clustering and Genetic Algorithm", personal correspondence.

[9] S.Mukherjee, "Order reduction of linear systems using eigen spectrum analysis", Journal of electrical engineering IE(I),Vol 77, pp 76-79,1996.

[10]G.Parmar, S.Mukherjee, R.Prasad, "System reduction using factor division algorithm and eigen spectrum analysis", Applied Mathematical Modelling, pp.2542-2552, Science direct, 2007.

[11]G.Pamar, S.Mukherjee, "Reduced order modeling of linear dynamic systems using Particle Swarm optimized eigen spectrum analysis", International journal of Computational and Mathematical Sciences, pp.45-52, 2007.

[12]Franklin, G.F.J.D.Powell, and M.L.Workman, "Digital control of Dyanmic systems", second, Addisan-Wesley 1990.

[13]Farsi M.,Warwick K. and Guilandoust M. "Stable reduced order models for discrete time systems", IEE Proceedings, Vol.133, pt.D, No.3,pp.137-141, May 1986.

[14]G.Saraswathi, "An extended method for order reduction of Large Scale Systems", Journal of Computing, Vol.3, Issue.4, April, 2011.

[15]G. Saraswathi, K.A. Gopala Rao and J. Amarnath, "A Mixed method for order reduction of interval systems having complex eigenvalues", International Journal of Engineering and Technology, Vol.2, No.4, pp.201-206, 2008.

[16]G.Saraswathi, "Some aspects of order reduction in large scale and uncertain systems", Ph.D. Thesis.