Summer 7-22-2012

Proceeding of International Conference on Software Technology and Computer Engineering STACE-2012

Prof. Dr. Srikanta Patnaik

Editorial

In the last century, most nations understood that science and technology was the best vehicle for economic upliftment. However, the problems that S&T was used to solve have been primarily local and customized to their needs. Today, the problems faced by nations are no longer, a concern of them alone. Humanity is devoting more and more attention to climate change, energy, water, disease, economic turbulence and terrorism etc., which are all of concern to the entire world and the solutions for which are beyond any individual nations or group of nations. If S&T has to provide upliftment of our humanity, if the research and development are in areas with porous borders between them and if the problems of the world know not of geographical borders, the education that becomes the foundation of all our science and technology, research and development also must of necessity become borderless. Computer science and information technology have already converged leading to Information and Communication Technology (ICT). Information Technology combined with bio-technology has led to bio-informatics. Similarly, Photonics is grown out from the labs to converge with classical Electronics and Microelectronics to bring in new high speed options in consumer products. Flexible and unbreakable displays using thin layer of film on transparent polymers have emerged as new symbols of entertainment and media tools. Now, Nano-technology has come in. It is the field of the future that will replace microelectronics and many fields with tremendous application potential in the areas of medicine, electronics and material science. I am sure about the use of nano-robot for drug delivery.

The rate at which new computer hardware products are arriving in the market is simply mind-boggling. As the technology advances, the size and the price of the devices come down, while the efficiency and capacity increase. The scenario is same in all cases, whether it is about internal components like processor, motherboard, RAM, graphics card, and hard disk or for peripheral accessories like mouse, keyboard, and monitors. Personal computers became popular only before about three decades back. But already there are huge piles of outdated and antique hardware components and devices. This is a tribute to the tremendous rate of development of latest technologies in computer hardware field. Perhaps, the newest entrant into the archeological catalogue of computer peripherals is CRT monitors. The sleek looking LCD monitors are spreading like computer virus. Data storage devices have attracted considerable attention of the technology developers. New kinds of storage devices such as newer versions of flash memory cards, hard disks using latest technology and disks of ever-increasing capacity are the results of advancement in latest technology in compute hardware. The memory size of the random access memory (RAM) cards is soaring to enable the smooth functioning of graphics animation software packages and streaming video websites. Also, computer motherboards have undergone substantial changes over the years. More and more functions are being added to the motherboard. Also, despite the incredible improvement in performance and functionalities, the price of these components has actually fallen steadily. The most vital component of a computer is the microprocessor. It is in this field that a battle of developing latest technologies in computer hardware takes place. The pace of development of microprocessor increases as the competition between the major processor chip manufacturing companies, Intel and AMD, intensifies. Both the companies are engaging in a neck and neck competition and continuously outdo each other in introducing new technologies.

The conference designed to stimulate the young minds including Research Scholars, Academicians, and Practitioners to contribute their ideas, thoughts and nobility in these disciplines of engineering. It's my pleasure to welcome all the participants, delegates and organizer to this international conference on behalf of IRNet family members. We received a great response from all parts of country and abroad for the presentation and publication in the proceeding of the conference. I sincerely thank all the authors for their invaluable contribution to this conference. I am indebted towards the reviewers and Board of Editors for their generous gifts of time, energy and effort.

Editor-in-Chief

Dr. Srikanta Patnaik
Chairman, I.I.M.T., Bhubaneswar
Intersceince Campus,
At/Po.: Kantabada, Via-Janla, Dist-Khurda
Bhubaneswar, Pin:752054. Orissa, INDIA.

An User-Oriented Efficient Economic Model for Cloud Computing

Gadaputi Mounika

CSE Department, Vignan University, Guntur,India
E-mail : gmounikacse13@gmail.com

Abstract **-** To design, develop and test a Cloud Based software application to automate the needs of Enterprise. The Cloud computing offers the availability and performance of services those users require to resolve their requests in reduced time and lower cost. In this solution, we will use IaaS software and we will create a web portal to manage the network resources. The software resources offered by our network will be based on ESX servers.

Keywords*—Datacenter,ESX server,Iaas,Cloudlet.*

I. INTRODUCTION

Today, applications are becoming more and more resource-intensive, either in memory or computing power. These resources are limited on a machine, the need to distribute applications over a network of computers, local or remote, is obvious. Grid computing is mainly focused on the provision of a significant amount of shared resources for intensive computing (research or physical applications). The protocols used are tailored to the specific topology of the network (low latency, high throughput). Cloud Computing is based instead on the provision of many services and data to users, without their having to manage the complex infrastructure required. Because of its support web services, it uses Internet protocols and is therefore subject to the constraints of the latter. The cloud computing environment suggests a future where we do not calculate on local computers, but on equipment centralized computing and storage operated by third parties. This environment has great flexibility and ease of use, availability of data and services became one of the biggest problems to address and improve.

With cloud computing, companies can raise large capacity in an instant without having to invest in new infrastructure, new staff or new software. This model of development was born after the giants of the Internet have created their own infrastructure to meet their own needs; they are now able to ensure their development, but also to share resources at competitive prices.

Cloud infrastructure modeling and simulation toolkits should provide support to economic entities such as brokers and exchanges Cloud for the negotiation of services between customers and suppliers. Finally, propose an economic model that brings together all beneficial criteria on the client side of the auction models.

II. CLOUD COMPUTING

Cloud computing can be defined as "*A type of parallel and distributed system consisting of a collection of interconnected computers and they are virtualized and dynamically generated and submitted as one or more computing resources based onservice level agreement established through negotiation between the provider and consumers*". Some examples of new infrastructure are Microsoft Azure Cloud Computing, Amazon EC2, Google and Aneka.

Computing power in cloud computing environments is provided by a collection of data centers, which are typically installed with hundreds of thousands of servers. The layered architecture of a typical cloud-based data centers is shown in Fig In the lower layers there are huge hardware resources (storage and application servers)which feed datacenters.

The servers are managed transparently by the level of virtualization, services and toolkits that enable the sharing of their capacity through virtual instances of servers. These virtual instances are isolated from each other, this helps to achieve fault tolerance and isolation of the security environment.

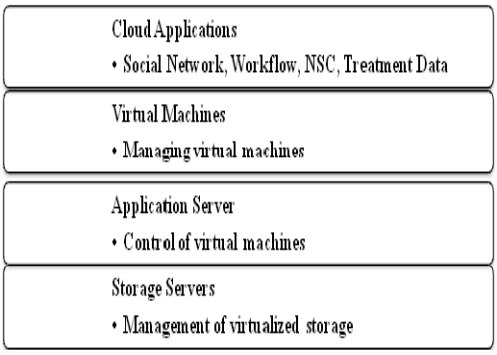


fig 1:Typical Datacenter

Cloud applications such as social networking, game portals, applications, economic and scientific workflows running in the highest layer of the architecture. The patterns of actual use of many real applications vary with time, mostly unpredictable ways. These applications have different requirements of Quality of Service (QoS) depending on the criticality of the time and modes of user interaction (online / offline).

Figure shows different technologies that have allowed the concept of cloud computing.

In the earlier days, two different technologies allowed to solve the problem of high computer resources demand: cluster computing and super computing. The first solution involved a computer network with different servers, and each user was redirected to one server. By contrast, the second solution referred to one powerful computer with specialized resources that was able to work for all users.

The term Grid Computing was used to describe a heterogeneous computer network where all devices were working together for a specific result. An example of this type of network is the project SETI@home, a collaborative software that uses wasted CPU cycles from personal computers, to analyze radio transmissions received from the space in search of intelligent life.

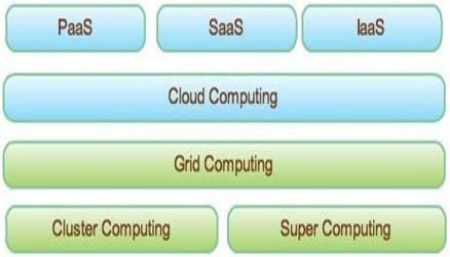


Fig: Cloud Computing Evolution

Although it is difficult to come up with a precise and comprehensive definition of Cloud Computing, at the heart of it is the idea that applications run somewhere on the "cloud". We do not know or care where exactly the application is.

This is not a big issue in a not critical scenario. As end users, we use web applications daily, without any concern about where the application actually runs. But on an enterprise scenario, where the data is confidential and must be carefully protected, this issue is critical.

The US NIST (National Institute of Standards and Technology) defines cloud computing as follows: "Cloud computing is a model for enabling convenient, on- demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model promotes availability and is composed of five essential characteristics (On-demand selfservice, broad network access, resource pooling, rapid elasticity and measured service), three service models (SaaS, PaaS and IaaS), and four deployment models (Private cloud, community cloud, public cloud and hybrid cloud)".

In Cloud Computing we can find mainly three different service models: Software as a Service (SaaS), Platform as a Service (PaaS) or Infrastructure as a Service (IaaS).

-**SaaS**: Delivery of an application as a service over the Internet, eliminating the need to install and run the application on the customer computer and simplifying the maintenance and support. An example of SaaS is Gmail, an e-mail application based on web browser developed by Google.

-**PaaS**: Delivery of a platform for deployment of applications without the cost and complexity of buying and managing the underlying hardware and software layers. An example of PaaS is Google App Engine, a platform for developing and hosting web applications in Google-managed data centers.

-**IaaS**: Delivery of a computer infrastructure, typically a platform virtualization environment as a service. Rather than purchasing servers, software, data center space or network equipment, clients rent those resources as a fully outsourced service. An example of IaaS is Amazon Elastic Compute Cloud (Amazon EC2) that allows users to rent virtual computers on which they can run their own computer applications. Also, Cloud Computing can be split into four different models depending on the deployment model of network:

-**Private cloud**: The cloud infrastructure is operated solely for an organization. It may be managed by the organization or by a trusted partner.In this type of clouds we cannot have infinite resources from a user point of view. To achieve this objective, we should have large free hardware resources that allowed absorbing work peak loads. On this hypothetical scenario we would have a high number of servers with a low utilization so we would not have any benefits in contrast to a scenario without cloud computing.

-**Community cloud**: The cloud infrastructure is shared by several organizations and supports a specific community that has shared concerns (e.g., mission, security requirements, policy, and compliance considerations). It may be managed by the organizations or by a trusted partner.

-**Public cloud**: The cloud is managed by a third-party provider, whereby resources are dynamically provisioned on a fine-grained, self-service basis over the Internet via web applications/web services. On this scenario we have to trust on this third-party provider that will manage our data to be safe and secure.

Moreover, a public cloud have elasticity. This means,that this type of network have infinite resourcesfrom a user point of view.

-**Hybrid cloud**: A hybrid cloud is a combination of two or more cloud models (private, community or public). On this scenario we can split risks, we can put the critical data on a private or community cloud and the not critical data on a public cloud. And we can have elasticity, if the private cloud is out of resources we can use the public cloud.

III. PROPOSEDSYSTEM

Cloudware offers enterprise-class managed solutions for business of all sizes. Utilizing a best-in-class architecture, IaaS Cloud allows its customers to focus on their core business while letting us provide them with an enterprise-level infrastructure for a fixed, predictable monthly fee.Our infrastructure utilizies best-in-class architectures and technologies from Tier 1 vendors, including, Vmware, and Netapp.

IV. PROBLEM STATEMENT

Life before cloud was really difficult if you want to start a business related to software then it was really problematic as you need one team just to maintain the whole system, one team for data security and also hardware maintenance team and for that you need lots of money and in current market software/hardware requirements are growing rapidly and today just because of rapid change of technology it is really difficult to compete with others. But today better alternatives are available as cloud computing is there. Cloud computing is internet based development and use of computer technology. With cloud computing you eliminate those headaches because you are not managing hardware or software that is the responsibility of the vendor like google/amazon/Microsoft/salesforce.com etc. These vendors provide storage. Processing power and computer application installed on their server you just need a computer connected with the net and according to your usage you have to pay.

Disadvantages:

- Single OS image per machine
- Software and hardware tightly Coupled.
- Running multiple applications on same machine often creates conflict
- Underutilized resources
- Inflexible and costly infrastructure

Advantages:

Cloud infrastructure Infrastructure-as-a-Service (IaaS) Cloud infrastructure services, also known as “Infrastructure as a Service (IaaS)”, delivers computer infrastructure - typically a platform virtualization environment - as a service. Rather than purchasing servers, software, data-center space or network equipment, clients instead buy those resources as a fully outsourced service. Suppliers typically build such services on a utility computing basis and amount of resources consumed (and therefore the cost) will typically reflect the level of activity. IaaS evolved from virtual private server offerings.

Thus Infrastructure-as-a-Service like Amazon Web Services provides virtual server instances with unique IP addresses and blocks of storage on demand. Customers use the provider's application program interface (API) to start, stop, access and configure their virtual servers and storage. In the enterprise, cloud computing allows a company to pay for only as much capacity as is needed, and bring more online as soon as required. Because this pay-for-what-you-use model resembles the way electricity, fuel and water are consumed; it's sometimes referred to as utility computing.Using the innovative Secured Cloud, you can:

- Provision customized cloud virtual machines on-demand
- Easily configure and horizontally scale servers to address traffic spikes
- Scale RAM allotment on the fly as need dictates

- Load balance your VMs and physical hardware with F5 Networks' technologies
- Turn your services on and off instantly to meet changing demands
- Manage everything through our custom-developed user interface
- Easily deploy a virtual firewall for an added layer of security
- Use APIs to efficiently integrate applications into the cloud

Intuitive User Interface:

In minutes, you can create a virtual machine (VM) in the Secured Cloud user interface (UI). Both highly functional and easy-to-use, our custom-developed UI makes setting up and managing VMs a simple exercise of point-and-click. There is no advanced technical knowledge required, no vast understanding of back-end systems needed, and no networking architecture expertise necessary. Just a simple and intuitive web interface to navigate.

V. CONCLUSION

The goal of cloud computing is to provide the applications necessary to operate a business, all on the Internet. The model developed in recent years has many advantages. The main aim being saving of permanent infrastructure that same company will not have to manage, whether at the level of maintenance and infrastructure costs (electricity, air conditioning ,troubleshooting, connectivity). The concept of machine no longer exists, the company stores its data and uses the cloud to work. The main objective of our work is to satisfy customers Clouds, while providing economic functions that reduce the cost of processing Cloudlets.

ACKNOWLEDGMENT

We are very thankful to the authors who have helps us to improve the quality and performance of our paper. We are grateful to our guide who guided us to our research work.

REFERENCES

[1] I. Foster, Y. Zhao, I. Raicu, and S. Lu, “Cloud Computing and Grid Computing 360-degree compared”, in Grid Computing Environments Workshop, pp.1–10, 2008.

[2] T. Rings, G. Caryer, J. R. Gallop, J. Grabowski, T. Kovacikova, S. Schulz, and I. Stokes-Rees, “Grid and Cloud Computing: Opportunities for Integration with the Next Generation Network”, Journal of Grid Computing, 7(3), pp.375-393, 2009.

[3] J. E. Smith, and R.Nair, “Virtual Machines: Versatile platforms for systems and processes”. Morgan Kauffmann, 2005.

[4] R. N. Calheiros, R. Ranjan, C. A. F. De Rose, and R. Buyya, “CloudSim: A Novel Framework for Modeling and Simulation of Cloud Computing Infrastructures and Services”, Technical Report, GRIDS-TR-2009-1, Grid Computing and Distributed Systems Laboratory, The University of Melbourne, Australia, 2009.

[5] F. Howell, and R. Mcnab, “SimJava: A discrete event simulation library for java”. In Proceedings of the first International Conference on Web-Based Modeling and Simulation, 1998.

[6] R. Buyya, C. S. Yeo, and S. Venugopal, “Market oriented cloud computing: Vision, hype, and reality for delivering IT services as computing utilities”. In Proceedings of the 10th IEEE International Conference on High Performance Computing and Communications, 2008

◈◈◈

A New Highly Secure and Efficient Routing Algorithm for Wireless Sensor Networks

Mohana [1] & N.K. Srinath [2]

[1]Dept. of CSE, R.V. college of Engineering, Bangalore-59, India
Email:mohana.rvce@gmail.com, mohana.@rvce.edu.in, srinath_nk@yahoo.com, srinathnk@rvce.edu.in

Abstract - Mobile Ad-hoc Networks (MANETs) comprises set of nodes connected by wireless links. The network is ad hoc because it does not rely on a preexisting infrastructure, such as routers in wired networks. Routing in MANETs is a challenging task due to dynamic topology and error prone shared environment.

Data is sent between nodes in a MANET by hopping through intermediate nodes, which must make decisions about where and how to route the data. MANET faces several problems because of node mobility, network traffic, network size, and the possibility of node faults. The efficiency and behavior of a MANET depends on how well information can be passed around and delivered.

In today's world the security vernubilities are increasing day by day. It is really difficult to route the packet with minimum packet loss. In this paper a new routing protocol is presented which would route the packets in a highly efficient way by introducing the concept of friend list, unauthenticated list and question mark list. The algorithm will avoid the malicious node by studying the network in an intelligent way. The proposed algorithm is also compared with other multipath routing algorithms namely Disjoint Multipath Routing (DMR), Trust based multipath routing (TMR), Message Trust based multipath Routing (MTMR) and the performance analysis proves that the proposed method will have better performance with respect to number of hops, route discovery time, packet loss, power consumed and energy.
The performance metric considered for proposed work are number of malicious nodes detected, number of hops, route discovery time, packet loss, power consumed and energy. The simulation results show that FACES protocol works much better and provides more security than the other multipath routing protocols.

Keywords - *Classes of Traffic, Data Rating, Dynamic Source Routing, Friend Rating, Net Rating, Route Discovery Time, Trust Level, Time To Live Period.*

I. INTRODUCTION

1. Disjoint Multipath Routing (DMR)

Initially secure connection has been established between source node to destination node. The DMR algorithm will find out the multiple routes from source to destination using DSR algorithm [1] [5]. After finding multiple routes, all the routes are sorted based on the route discovery time. Then it will choose the best four routes which are having minimum time delay. In this method the message is split into parts. This protocol takes advantage of the shortest path between the source node to destination node. Then it routes the four encrypted parts through four different routes. In this method to decrypt the original message all the encrypted parts are required. The security of this method lies in the fact that enemy node needs all the encrypted parts to decrypt the original message.

II. TRUST BASED MULTIPATH ROUTING (TMR)

TMR provides a method of message security using trust based multipath routing. In this approach, less trusted nodes are given are not given the encrypted parts of a message, thereby making it difficult for malicious nodes to gain access to the minimum information required to break through the encryption strategy [2]. Using trust levels, it makes multipath routing flexible enough to be usable in networks with "vital" nodes and absence of necessary redundancy. In addition, using trust levels, it avoids the non trusted nodes in the routes that may use brute force attacks and may decrypt messages if enough parts of the message are available to them.

Secure connection has been established between source node to destination node. The TMR algorithm will find out the multiple routes from source to destination using DSR algorithm. After finding multiple

routes, all the routes are sorted based on the trust level. Then it will choose the best route which is having maximum trust level. In this method the message is split into parts. Then it routes the encrypted parts through best single route. The following table gives the description about the trust levels and the trust levels are varied between -1 to 4 [2] [3].

Sl No	Trust Value	Meaning	Description
1	-1	Distrust	Completely untrustworthy.
2	0	Ignorance	Cannot make trust-related judgment about entity.
3	1	Minimal	Lowest possible trust.
4	2	Average	Mean trustworthiness.
5	3	Good	More trustworthy than most entities.
6	4	Complete	Completely trust this entity.

Table1: Trust levels.

III. MTMR ROUTING ALGORITHM

MTMR uses a trust assignment and updating strategy which can be used to identify and isolate malicious nodes. The MTMR algorithm will find the routes by using the DSR algorithm. If the route has been fined for the first time then the MTMR algorithm will have the all routes obtained from DSR with trust level of zero. The MTMR algorithm will then choose a route with minimum time delay. Then if the route contains the malicious node then the trust level of the node is decremented otherwise the trust levels of all the nodes in the best route will be incremented [4].

Unlike TMR MTMR routing algorithm does not assign random trust levels instead the trust levels are assigned only to those nodes which behave properly and deliver the packets successfully.

IV. FRIEND BASED AD-HOC ROUTING (FACES)

This protocol is used to find out the secure route from source to destination. It will route the packets in a highly efficient way by introducing the concept of friend list, unauthenticated list and question mark list. The algorithm will avoid the malicious node by studying the network in an intelligent way. If any malicious node will come in the best route, it will detect that node and it will put it in the question mark list.

Source node will pick the intermediate node from the friend list. If the Friend List is empty then the node will obtain the unauthenticated list and pick one of the node as intermediate node. The value of TTL is decremented each time intermediate node is picked. Once the TTL becomes zero we have to use min hop routing so that the packet can be delivered to the destination at the faster rate.

The new routing algorithm will make use of following parameters

Question Mark List (QML): The list of nodes which are deemed suspicious by a particular node. This list is stored for each and every node in its data structure.

Unauthenticated List (UL): The list of nodes of which no security information is present.

Friend List (FL): This is the list of nodes which convey trust. Like the question mark list, a friend list is also stored for each node in its data structure. Friends are rated on a scale of 0 to 10.

FREQ: Friend Sharing Request, this is a control packet which is used to initiate friend sharing. A node receiving this packet replies with the nodes in its friend list, unauthenticated list and the question mark list.

DR: Data Rating, this is the rating given to nodes after they transmit some amount of data for the source node.

FR: Friend Rating, this is the rating computed when nodes share their friend lists.

NR: Net Rating, this rating is computed as a weighted mean of DR and FR.

OR: Obtained Rating, rating received during the friend sharing stage.

4.1 SHARE FRIEND STAGE ALGORITHM

This is the stage in which a node will exchange the friend list with other node in the network

The following figure (1) gives brief information about the share friend stage for various cases between two nodes namely A and B.

Friend initiator NodeA	Friend Giver NodeB	Common Nodes	Uncommon Nodes	Nothing
FL=[ ]	FL=[ ]			FR=FR+1 DR=DR+1 NR=NR+1
FL=[ ]	FL !=[ ]	FR=FRnodeB+1 NR=NRnodeB+1 DR=DRnodeB+1	FR=FR+1 DR=DR+1 NR=NR+1	
FL!=[]	FL=[]	FR=FRnodeA+1 NR=NRnodeA+1 DR=DRnodeA+1	FR=FR+1 DR=DR+1 NR=NR+1	
FL!=[]	FL !=[]	FR=FRnodeA+ FRnodeB +1 NR=NRnodeA+ FRnodeB +1 DR=DRnodeA+ FRnodeB +1	FR=FR+1 DR=DR+1 NR=NR+1	

Figure1:Friend sharing stage

Friend sharing is a periodic process which is chiefly responsible for the security of the algorithm. To accomplish friend sharing we use the control packet *FREQ* (Friend sharing request). The node receiving the *FREQ* replies with the nodes in its friend list, unauthenticated list and the question mark list. The rules for friend sharing are as follows:

1. Any node can ask for a friend sharing request.
2. After friend sharing, challenges are initiated for those nodes which were not in the friend list.
3. If a node is already in the friend list the node updates its friend list.

Let us consider that the node *A* shares a list with node *B*. Then the friend sharing process is carried out as follows:

STEP 1: As the network is initialized each node starts the friend sharing process, which leads up to a challenge to start with the formation of friend list.

STEP 2: If node *B* is in the list of node *A*, then if a particular friend of node *B* is not present in the list of node *A*, node *A* includes it in its list and initializes the Friend Rating as the Obtained Rating from *B* and the Data Rating to Zero. Net Rating is calculated on the basis of pre determined weights.

STEP 3: if the node B is in the list of node A and if a

Particular friend of node *B* is present in the list of node *A* then the Friend Rating is calculated.

Data Rating: The data rating is updated by a node for its friend on the basis of amount of data it transfers for it. The DR of a friend node varies according to the number of data packets transferred through it. The net DR is calculated as a moving average of the last five data ratings. Equation (1) describes the moving average relation between a data rating *i and* the previous five data ratings:

$$DR(i) = \frac{DR(i-1) + DR(i-2) + DR(i-3) + DR(i-4) + DR(i-5)}{5} \quad \ldots\ldots (1)$$

Friend Rating: During the *Friend Sharing* stage a node *A* asks for the friend list of node *B* and incorporates the rating of friends in the following way:

1. If the node *A* and node *B* have a common friend *C*, *t*hen node *A* obtains the rating of node *C* from the node *B* as:

$$obtaine\ d\ rating = \frac{Net\ \ rating\ \ of\ \ B\ in\ \ list\ \ ofA * Net\ \ rating\ \ of\ C\ in\ list\ \ of\ \ B}{10} \quad \ldots\ldots. (2)$$

The idea behind equation (2) is *to* incorporate the trust that node *A* has on node *B* while obtaining the rating of node *C* from it.

2. After the Friend Sharing Process has finished, each node adds up the *OR* from various nodes and divides them with the sum of the rating of those nodes from which it obtained the *OR*. In other words, the *FR* is the weighted average of the Net Ratings obtained during the Friend Sharing stage, where the weights are basically the Net rating of the friend responding to a particular Friend Sharing request.

Net Rating: The idea behind calculating DR and FR is to have two opinions in front of each node. This is done because malicious nodes can identify some nodes for which they would work properly while for some they would drop packets. The DR acts as the soul opinion of the host node and FR acts as the opinion of its friend nodes. The Net Rating (NR) would be a weighted mean of the two ratings as given in equation (3):

$$NR = \frac{W_1 * DR + W_2 * FR}{W_1 + W_2} \quad \ldots\ldots\ldots\ldots (3)$$

Where *W1* and *W2* would be the weights assigned to DR and FR respectively. The values of *W1*and *W2* are network dependant and can be learnt with experience.

4.2 Form Un-Authenticated List

The nodes will find out the nodes in the transmission range or within the coverage area and then it maintains a list of nodes which are reachable by the nodes on its data structure.

4.3 Sequential challenges

When a source discovers that the data was not transmitted properly it initiates a sequential challenge. To explain it let us consider the path as, $S \rightarrow A \rightarrow B \rightarrow C \rightarrow D$ where S and D are the source and the destination nodes and A, B, C are the intermediate nodes. On discovering data transmission problem after waiting for a back off interval the following process takes place.

Step1.The source challenges node **A**. If it is not able to complete the challenge it removes it from the friend list and places it in the question mark list. It then tries to route the data through the next best path.

Step 2. If **A** successfully completes the challenge, it challenges its neighbor and reports the result backward To **A.** Node **A** then takes action similar to Step 1.

Step 3. Similarly, if the node **B** is also authenticated, it challenges its neighbor node **C** and reports backwards to node **A**.

Step 4. If every node is authenticated then node **D** gets to know the result of the sequential authentication, when node **C** tries to authenticate it, it responds to it and sends a backward message attaching the number of data packets received by it. Since everyone is authenticated no one lies, and attaches the corresponding packets truthfully. The source thus comes to know where the packet drops has taken place.

We emphasize here that during the sequential challenge process no node would be able to detect whether it is a sequential challenge or a regular friend sharing process till it is authenticated. And thus a malicious node would be caught in the process and eliminated from the list of trusted node.

4.4 FRIEND ROUTING PROTOCOL

The Friend routing protocol will perform the following steps

Step 1.The Source Node will first find the set of intermediate nodes by doing a lookup in its Friend List.

Step 2.If the friend list is empty then the source node will look into the unauthenticated List.

Step 3. If the unauthenticated list is empty the friend list has no other choice of picking the node from question mark list.

Step 4. The source node will check whether it contains the destination node in its list if yes then the faces protocol will transmit the data directly to the intermediate node.

Step 5.The intermediate node will then become the source node (picked up during either the steps 1 2 and 3).

Step 6.The process repeats until the Time to live period expires or destination node is reached.

Step 7. If the TTL =0 then the current node will always pick a node which is closer to destination so that the destination can be reached at a faster rate.

V. SIMULATION ANALYSIS

The simulations have been performed using IDE, java eclipse Galileo. Database used My SQL and some of the software packages used are Jdk 1.6, Jre6 and Jar files of Structs framework. Simulation results are exported using Toad software and performance graphs are plotted using MATLAB.

Source Node	Destination Node	Coverage Area
5	45	30

Table 2: Inputs to Routing Algorithms

1. DMR Algorithm Output

Output of Stage1 for DMR Algorithm

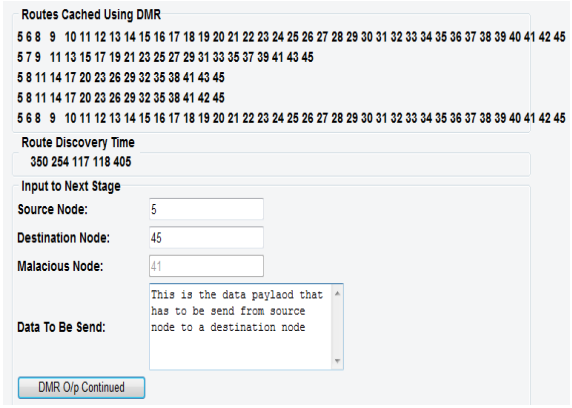


Figure 2: Multiple Routes Discovered using DSR, data payload to be send**.**

Figure 2 shows the multiple routes that have been discovered from source node to a destination node using DSR (Dynamic Source Routing) algorithm and there corresponding Route Discovery time as well. The user is also entering the data payload that has to be sent from source node to destination node.

Output of DMR Algorithm

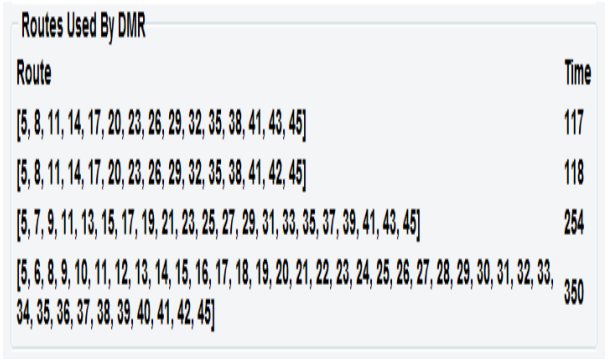


Figure3: Routes Chosen by DMR

Figure-3 shows the multiple routes that are discovered using DMR algorithm from source node to the destination node. DMR will select best four routes which is having a less route discovery time.

PACKETS

Packet1:	5	45	This is the data pay	1
Packet2:	5	45	laod th	2
Packet3:	5	45	at has to be s	3
Packet4:	5	45	end from source node to a destination node	4

ENCRIPTED PACKETS

Encripted Packet1:	5	45	[B@1fcd402	1
Encripted Packet2:	5	45	[B@1c2ec05	2
Encripted Packet3:	5	45	[B@1558dc	3
Encripted Packet4:	5	45	[B@17d03c5	4

Figure4: Packet Formation Output

Figure4 shows the Packet Formed using the Triple Des Encryption for the data fragments. These data fragments would be sent over multiple independent routes from source node to destination node.

2. TMR Algorithm Output

Trust Level	Route
0	5 8 11 14 17 20 23 26 29 32 35 38 41 42 45
-2	5 6 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 45
-1	5 8 11 14 17 20 23 26 29 32 35 38 41 43 45

Best Route of TMR

5 8 11 14 17 20 23 26 29 32 35 38 41 42 45

TRUST LEVEL 0

Figure-5: TMR Algorithm Output

Figure-5 shows the TMR algorithm having multiple routes from source node to the destination node. The TMR algorithm will choose a route which is having the maximum Trust from Source Node to Destination node in the network. The encrypted data fragments will be sent in the single best route.

3. MTMR Algorithm Output

MTMR Routing Algorithm takes an additional parameter as input i.e threshold trust of a route.Threshold Trust=40

Route Using MTMR AND TRUST MAP

TRUST 152 ROUTE [5, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 45]

TRUST 77 ROUTE [5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45]

TRUST 56 ROUTE [5, 8, 11, 14, 17, 20, 23, 26, 29, 32, 35, 38, 41, 43, 45]

TRUST 60 ROUTE [5, 8, 11, 14, 17, 20, 23, 26, 29, 32, 35, 38, 41, 42, 45]

Best Route Possible At this Time

5 6 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 45

POSSIBLE TRUST LEVEL 152

Route Using MTMR

5 6 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 45

TRUST LEVEL Using MTMR 152

Figure-6: M TMR Algorithm Output

Figure-6 shows the MTMR algorithm having multiple routes from source node to the destination node. The MTMR algorithm will choose a route which is having the maximum Trust from Source Node to Destination node in the network. The additional thing happening in MTMR is the nodes which are in the best route will have their corresponding trust levels incremented by a factor of 1.

4. Friend based Ad-hoc Routing output

The Friend will also take TTL has an additional input parameter as compared to other algorithms.

Net Rating	Route
118.0	5 5 8 11 14 17 20 23 26 29 32 35 38 41 44 45
106.0	5 5 8 11 14 17 20 23 26 29 32 35 38 41 44 45
124.0	5 5 8 11 14 17 20 23 26 29 32 35 38 41 44 45
112.0	5 5 8 11 14 17 20 23 26 29 32 35 38 41 44 45

Best Route Discovered and Rating

5 8 11 14 17
20 23 26 29
32 35 38 41
44 45

Rating Route

124.0 5

Figure-7: Friend Routing Algorithm Output

Figure-7 shows the Friend Based Routing Algorithm output .As seen from the figure the Friend Routing Protocol has discovered all the routes by picking based on the combination of friend rating, data rating and net rating. The Friend Routing has chosen the route which is having the maximum rating as the best route.

Share Friend Stage Output

NodeId	Friend Rating From Giver	Friend Rating From Initiator	Data Rating From Giver	Data Rating From Initiator	Net Rating From Giver	Net Rating From Initiator
6	1	1	1	1	1	1
9	1	1	1	1	1	1
12	3	3	3	3	6	6
32	4	0	3	0	6	0
35	4	4	3	3	6	6
38	4	0	3	0	6	0
45	4	0	3	0	6	0
41	4	0	3	0	6	0
44	4	0	3	0	6	0
30	10	10	11	11	65	65
25	10	18	11	19	65	101
26	10	10	15	11	108	65
29	10	10	15	11	108	65
23	10	10	27	23	270	211
20	10	10	27	23	270	211
17	10	10	27	23	270	211
14	10	10	27	23	270	211
11	10	10	29	25	302	240
8	10	10	29	25	302	240
5	10	10	41	33	520	369

Figure-8: Friend Sharing Stage Output

Figure 8 shows the output of Friend Sharing Stage As seen in the figure the friend list of Friend Stage

Initiator is shown where the Node Id are friend node ids , Friend Rating from giver is rating allocated from node 5. Friend Rating from initiator is as per node 4. Similarly Data Rating and Net Rating are shared between two nodes Node 4 and Node5.

VI. PERFORMANCE ANALYSIS OF ALGORITHMS

1. No of Hops

Figure 9 shows the number of hops taken from source to destination for all four algorithms DMR, TMR, MTMR and FACES. DMR will take more number of hops.TMR and MTMR will take almost equal number of hops. FACES will take less number of hops. From the figure conclude that FACES algorithm is the best w. r. t number of hops.

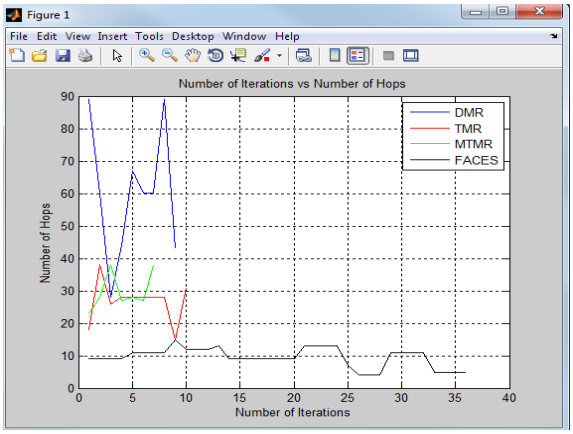


Figure-9: Number of hops

2. Route Discovery Time

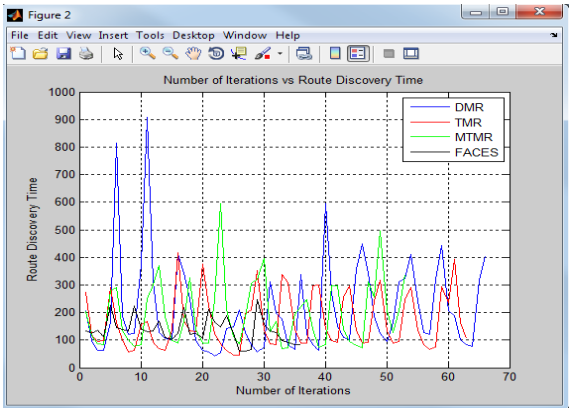


Figure-10: Route Discovery Time

Figure-10 shows the route discovery time taken for all the routes from source to destination for all four algorithms. DMR and MTMR are taking more time to establish path between source node to destination node. TMR is taking medium route discovery time. From the figure conclude that the route discovery time taken by FACES algorithm is less as compared to DMR, TMR and MTMR

3. Packet Loss

Figure 11 shows the packet loss taken for all the routes from source to destination for all four algorithms. DMR is having a more number of packet losses. Packet drop is minimal in FACES, as it will detect more malicious nodes and efficiently discards routes containing malicious nodes. But other multipath routing protocols drop a larger number of packets as they route through a greater number of nodes and thus increasing the chances of routing data through malicious nodes. From the figure conclude that FACES algorithm is the best when compared to DMR, TMR, and MTMR

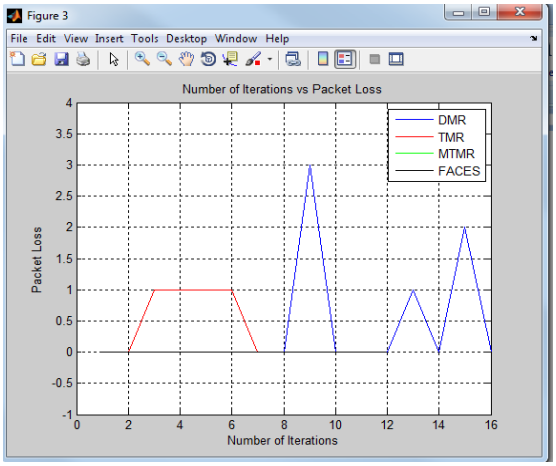


Figure-11: Packet Loss

5. Power consumption

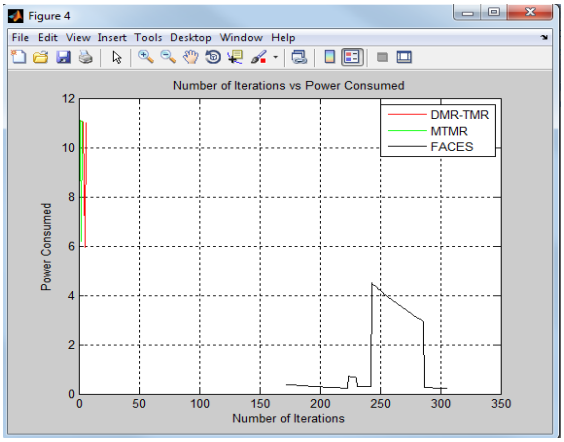


Figure-12: Power consumption

Figure-12 shows FACES use a minimum power as compared to other trust based multipath protocols. From the figure conclude that power consumed during the

route discovery mechanism in FACES algorithm is less when compared to other algorithms.

6. Energy

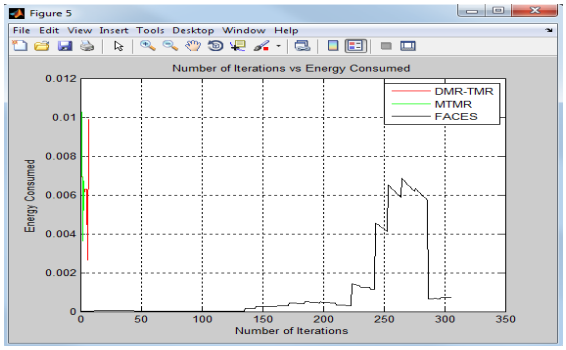


Figure-13: Energy consumed**.**

Figure-13 shows FACES use a minimum energy as compared to other trust based multipath protocols. This is due to the fact that, other protocol uses promiscuous mode which use much more energy as all nodes have to overhear all the transmissions. From the figure conclude that energy consumed in FACES algorithm is less when compared to other algorithms.

VII. CONCLUSION AND FUTURE WORK

Mobile Ad-hoc network (MANETs) due to its dynamic nature has many challenges. Some of the major challenges are number of malicious nodes detected, number of hops, route discovery time, packet loss, power consumption and energy

Many Routing algorithms namely DMR, TMR, MTMR and FACES have their own way in order to establish the trust and transmit packet securely. But Friend based protocol proved to be best in terms of number of malicious nodes detected, number of hops, route discovery time, packet loss, power consumption and energy.

In future we plan to implement the existing secure routing protocols such as the ARIADNE and ARAN and compare them with the FACES protocol. The system handles only text as message for data packets does not handle multimedia data packets. In future it can be enhanced to include multimedia packets.

REFERENCES

[1] David B. Johnson David A. Maltz Josh Broch "DSR: The Dynamic Source Routing Protocol for Multi-Hop Wireless Ad Hoc Networks" *Computer Science Department, Carnegie Mellon University*.

[2] Alfarez Abdul-Rahman & Stephen Hailes "A Distributed Trust Model" *Department of Computer Science, University College London, Gower Street, London WC1E 6BT, United Kingdom*.

[3] A.A. PIRZADA and C. McDONALD "Trust Establishment in Pure Ad-hoc Networks" *School of Computer Science & Software Engineering, The University of Western Australia, 35 Stirling Highway, Crawley, W.A. 6009, Australia*, Wireless Personal Communications (2006) *37: 139–163, DOI: 10.1007/s11277-006-1574-5 C _ Springer 2006*.

[4] Su Bing, Ma Zheng-Hua,Sun Yu-Qiang"A Trusted-Based Encryption Mechanism for Efficient Communication Over Wireless Network" Wireless Communications, Networking and Mobile Computing, 2008. WiCOM '08 4th International Conference, *Digital Object Identifier: 10.1109/WiCom.2008.1111*

[5] Al-Mekhlafi Z.G, Hassan R, "Evaluation study on routing information protocol and dynamic source routing in Ad-Hoc network", *Information Technology in Asia (CITA 11), 2011 7th International Conference, E-ISBN : 978-1-61284-130-4, INSPEC Accession Number: 12208218 Digital Object Identifier :10.1109/CITA.2011.5999535 Issue Date : 30 August 2011, pp 1-4.*

[6] L.Wang and N.-T. Zhang, "Locally forwarding management in ad-hoc networks," in *Proc. IEEE Int. Conf. Communications, Circuits and Systems and West Sino Expositions, Jun./Jul. 2002, pp. 160–164*.

[7] D. Johnson and D. Maltz, "Dynamic source routing in ad hoc wireless networks," in *Book Chapter inMobile Computing*, T. Imielinski andH. Korth, Eds. Dordrecht, The Netherlands: Kluwer, 1996, pp. 131–181.

[8] A. Wood and J. A. Stankovic, "A taxonomy for denial-of-service attacks in wireless sensor networks," in *Handbook of Sensor Networks: Compact Wireless and Wired Sensing Systems*. Boca Raton, FL: CRC, 2005, pp. 32:1–32:20.

[9] W. Diffie and M. E. Hellman, "New directions in cryptography," *IEEE Trans. Inform. Theory*, vol. IT-22, no. 6, pp. 644–654, 1976.

[10] M. S. Obaidat and N. Boudriga, *Security of e-Systems and Computer Networks*. Cambridge, U.K.: Cambridge Univ. Press, 2007.

[11] K. Sanzgiri, B. N. Levine, C. Shields, B. Dahill, and E. M. Belding- Royer, "A secure routing protocol for ad hoc networks," in *Proc. 10th IEEE*

Int. Conf. Network Protocols (ICNP), Paris, France, Nov. 12–15, 2002, pp. 78–89.

[12] Y. Hu, A. Perrig, and D. B. Johnson, "Ariadne: A secure on-demand routing protocol for ad hoc networks,"*Wireless Netw.*, vol. 11, no. 1–2, pp. 21–38, Jan. 2005.

[13] S. Marti, T. J. Giuli, K. Lai, and M. Baker, "Mitigating routing misbehavior in mobile ad hoc networks," in *Proc. MobiCom 2000*, Boston, MA, Aug. 2000, pp. 255–265

◈◈◈

Implementation of New Corporate Decision Control Paradigm with Android

[#]B.J.D Kalyani & [*]K.V.D Prasad

[#]Department of Information Technology, Swarna Bharathi Institute of Science & Technology, Khammam,A.P,India
[*]Department of Customer Care , Sahara India company Ltd., Khammam, A.P. India
E-mail : kjd_kalyani@yahoo.co.in & bjd.kalyani@gmail.com

Abstract - This paper proposes a framework for business oriented self configurable visual indicators to control business navigation for example instantaneous meter to indicate today's sales up to this minute, Alarm blinker indicator to show stocks exceeding desired inventory levels etc using a mobile platform. Android is selected as the platform for this Decision control System navigator panel. All business controls and Indicators are to be developed as generic Objects. Business users will have a choice to select the type of Control/Indicator they need and set their behavior (like intervals for scanning, upper lower limits etc.). The Android mobile shall connect to the corporate data warehouse like SAP or ORACLE APPLICATION Server to fetch information from time to time and provide visual clues to the corporate business user for necessary action. Mobile computing has come of age pushing back laptops and desktops by corporate users. Android mobile platform pioneered by Google internet application framework is projected to be the future of mobile computing. This paper proposes a new decision control system paradigm to assist corporate managers on a generic mobile platform using Android.

Keywords- *Android, Business Intelligence (BI), On Line Analitical Processing(OLAP),Software Development Kit(SDK),Mobile Digital Dash Boards,Enterprise Resource Planning,Customer Relationship Management.*

I. INTRODUCTION

Mobile Digital dashboards allow managers to monitor the contribution of the various departments in their organization. To gauge exactly how well an organization is performing overall, Mobile Digital Dashboards allow you to capture and report specific data points from each department within the organization, thus providing a "snapshot" of performance.

The **Open Handset Alliance [1]**, a group of more than 30 technology and mobile companies, is developing Android: the first complete, open, and free mobile platform. Android [2] is a software stack for mobile devices that includes an operating system, middleware and key applications. This current version of the Android SDK [3] provides the tools and APIs necessary to begin developing applications on the Android platform using the Java programming language. Developers have full access to the same framework APIs used by the core applications. The application architecture is designed to simplify the reuse of components; any application can publish its capabilities and any other application may then make use of those capabilities subject to security constraints [4] enforced by the framework. This same mechanism allows components to be replaced by the user.

While a driver driving a vehicle on the road has several controls like Blinkers (to show direction), Speedometer (to indicate current speed of the vehicle), fuel meter (to indicate current level of fuel), accelerator and break (to control the speed) steering (to change direction) etc. the tools available to corporate managers to navigate in a more complex business space are limited.

Android Mobile Decision control paradigm gives mobile workers the ability to make information-driven decisions, regardless of location, by securely receiving and interacting with Enterprise Resource Planning Systems like SAP, Oracle App etc. The Situation Aware mobile widgets stored in the Android mobile platform shall access and display relevant information to the mobile user.

Benefits of using DSS of mobile digital dash board include:

1. Make the problems, abnormalities, or difference from standards visible and thus corrective action can be taken immediately.
2. Helps the corporate managers to monitor what's going on at a quick glance
3. Visual appearance of performance measures
4. Ability to identify and accurate negative trends
5. Helps the corporate managers to take corrective action as close to the time of occurrence of the problem as possible
6. provide immediate feedback to people
7. Measure efficiencies/inefficiencies
8. Ability to generate detailed reports showing new trends
9. Ability to make more informed decisions based on collected business intelligence
10. Align strategies and organizational goals
11. Save time over running multiple reports
12. Gain total visibility of all systems instantly

II. KNOWLEDGE DISCOVERY USING BUSINESS INTELLIGENCE (BI)

Business Intelligence is the term most widely used in commercial organizations and is generally refers to the increased utilization of various forms of information technology (IT) to capture, store, extract, manipulate, analyze and communicate data and information of all forms by firms across industry sectors. There is a greater accessibility to increased amounts of information around the globe than any time in the past. As a result firms can better transform vast amounts of data into a more vital asset, information that ultimately enhances the knowledge level of individuals across functional areas of an organization.

Business Intelligence of given organization is greatly depends on proper utilization of information technology. Uncertainty of business issues that really affect day to day operations at the firm level by enhanced business intelligence Firms of all sizes and industry types are utilizing these technologies to help augment their operations to compete, survive and thrive in this new dynamic economy.

Efficiency of firms can be increased by implementing state-of-art IT in knowledge discovery through data mining and business intelligence in describing the process. More specifically, it focuses on the high-end analytical software technologies like data mining and how this technology, along with other applications such as On Line Analytical Processing (OLAP), can help decision makers extract information and knowledge from the vast amounts of data they collect on a day by day and minute by minute basis. This work addresses the issues of data mining in an e-commerce environment as well, connecting the more traditional "brick and mortar" firm structure to the growing "click and mortar" enterprise.

Elaborate IT networks enable users to extract data (demographic and transactional) into structured reports, which can be distributed throughout an enterprise via intranets. As a result, information corresponding to particular functional areas specifically Value added information as well as aggregate information is more readily available to consumers of the data.

Example Scenarios

1. An executive of a courier company want to know from time to know the courier bookings of the day at various collection points. He configured an Android mobile widget to connect to the central transaction server and fetch the required information.

Figure1: Ex1 of Decision Control System Paradigm

When he selects the location the widget displays a dash board indicating that the current level of booking is 136 against a max possible booking for the day as 500 shown in figure1.

2. A bank manager in a corporate bank is interested in knowing the money dispersed and balance in hand to monitor liquidity. He can configure an Android dash board widget to do that as shown in figure2.

Figure2: Ex2 of Decision Control System Paradigm

III. BI WITH MOBILE COMPUTING

The emergence of the extended enterprise has changed the dynamics of business information delivery to the point where firms are demanding to extend the value of existing Business Intelligence (BI) investments to provide actionable information to a distributed or mobile user community [5]. Delivering this information to a mobile workforce via the new generation of mobile computing devices is becoming a requirement of the mobile work force[6]. The mobile user is fundamentally different from the stationary user in the following ways:

1. The mobile user is moving, at least occasionally, between known or unknown locations.
2. The mobile user is typically not focused on the computing task.
3. The mobile user frequently requires high degrees of immediacy and responsiveness from the system.
4. The mobile user is changing tasks frequently and/or abruptly.
5. The mobile user may require access to the system anywhere and at any time.

IV. SITUATION AWARENESS (SA) WITH MOBILE DIGITAL DASH BOARDS

The idea of Mobile Digital Dashboards followed the study of decision support systems in the 1970s. Many systems were developed in-house by organizations to consolidate and display data already being gathered in various information systems throughout the organization. Today, Digital dashboard technology [7] is available "out-of-the-box" from many software providers. Some companies however continue to do in-house development and maintenance of dashboard applications. For example, GE Aviation has developed a proprietary software/portal called “Digital Cockpit" to monitor the trends in aircraft spare parts business. Specialized dashboards may track all corporate functions. Examples include human resources, recruiting, sales, operations, security, information technology, project management, customer relationship management and many more departmental dashboards. In order to assist executives in making complex decisions, a new kind of computer system has been developed - the executive dashboard. Executive dashboards are systems that allow company executives to view key business facts, providing a circumspective view that is designed to support effective decision-making.

V. SYSTEM SPECIFICATIONS

I) Business user specification:

Table1: Android Phone HTC Dream Specifications

Manufacturer	HTC
Model	HTC dream
Carrier	T-Mobile[8]
Screen	3.2 in (81 mm) HVGA (480×320) (180 ppi) 65K color capacitive touch screen
Camera	3.2 megapixel with auto focus
Operating system	Android 1.5[9]
Input	Capacitive touch screen, sliding QWERTY Keyboard, Trackball
CPU	Qualcomm MSM7201A ARM11 @ 528MHz
Default ringtone	G1 Mix tape
Memory	192 MB DDR SDRAM 256 MB Flash
Memory card	microSD
Networks	Quad band GSM / GPRS / EDGE: GSM 850 / 900 / 1800 / 1900
	Dual band UMTS / HSDPA / HSUPA: UMTS 1700 / 2100 (US/Europe) (7.2/2 Mbit/s)
Connectivity	Bluetooth 2.0, IEEE 802.11 b/g, ExtUSB
Battery	1150 mAh
Physical size	117.7 mm x 55.7 mm x 17.1 mm (4.60 in x 2.16 in x 0.62 in)
Weight	158g w/ battery
Series	A Series
Successor	HTC Magic
Hearing Aid Compatibility	(none)

II) Administrator Specifications:

Table2: Specifications of the Server

Manufacturer	IBM
MPN	7025-F50
Key Features	
Form Factor	Tower
Processor	PowerPC 604E332 MHZ
Processors Qty.	1
Installed Memory	128MB(SDRAM)

Compatible OS Platform	IBM AIX 4.3 Unix
PROCESSOR	
Processor Manufacturer	IBM
Processor Type	PowerPC 604E
Processor Speed	332 MHZ
Installed Qty.	1
Max Supported Qty	4
Processor Upgradeability	Upgradeable
MEMORY	
Installed RAM	128 MB
RAM Technology	SDRAM
Max RAM	3 GB
Hard Drive Capacity	9.1 GB
Hard Drive Interface	SCSI-2 Fast Wide(16-bit)
NETWORKING	
Networking Type	Network Adapter-Ethernet
MOTHER BOARD	256 KB

VI. DESICION CONTROL SYSTEM PARADIGM

Decision control System using Android is a Component Based Software Development approach that uses bottom up approach of development.

I) System Architecture:

The transaction information of the corporate world is locked into ERP systems like SAP, Oracle Apps. Etc. This information can be rationalized, translated and stored periodically in to Business Intelligence Warehouses [10]. Using multi-dimensional modeling techniques, the transaction information is transformed into Decision Control System information. It could be summarized and stored as HYPERCUBES in the data warehouse. The mobile devices and remote user community normally will access these information chunks stored in the Hypercube. The Situation Aware mobile widgets stored in the Android mobile platform shall access and display relevant information to the mobile user as shown in the figure3.

The Administrator uses the SERVER system with desktop applications to do these functions. Most of these functions are performed interactively using tools provided by the CRM Vendor via SAP in the present context and he uses the specifications in Table2. SAP Net Weaver Mobile is the foundation of SAP mobile application suite called SAP app.'s for Mobile Business .The platform is also intended for enterprise and partners to develop custom mobile applications based on SAP enterprise applications such as SAP ERP, SAP CRM as well as non-SAP enterprise applications.

It includes graphical tools to model, develop and deploy powerful mobile composite applications, an integration framework to connect to enterprise software, and the resources required to manage and monitor the life cycle of composite applications [11]. At the heart of SAP Net Weaver Mobile is the data orchestration engine. It determines what pieces of data from SAP and non-SAP business applications are needed by each mobile user and distributes them based on distribution rules modeled on the unique business rules of the individual organization [12].

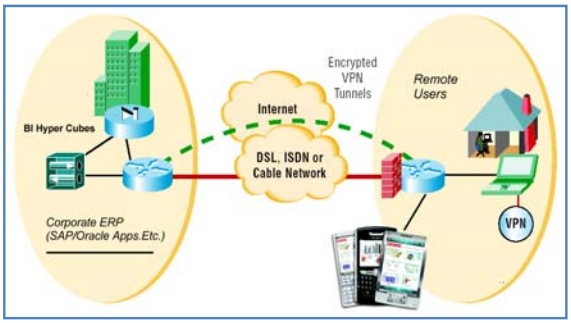


Figure 3: Decision Control Paradigm Architecture

By leveraging an event-driven architecture, back-end enterprise applications can publish to the distribution engine. The engine then routes and transmits the changes to all subscribing mobile devices as the devices connect. Within the mobile middleware it is possible to model data objects in order to store the data from the backend. It is also possible to model rules and dependency in order to distribute the data to the corresponding devices as shown in figure4.

The client side application development offers a integrated toolset to maximize the design of a sophisticated mobile application and to minimize the manual coding steps. This leads to an increase of software quality and developer productivity. The mobile client-side application development toolset is a plug-in for the Net Weaver Developer Studio (NWDS), which is an Eclipse based development environment [13] with graphical tool support.

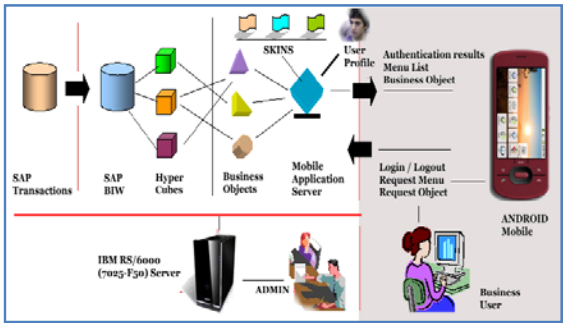


Figure4: System Flow Diagram

SAP NetWeaver Mobile includes device management and administration functionality that enables enterprises to manage key aspects of their composite application deployment. Composite applications can be deployed automatically without requiring user involvement, which reduces the cost of administration. Mobile devices can be managed centrally by user profile, group, or individual user. The status of installations and the vital signs of mobile devices can be monitored and diagnosed remotely, and mobile device settings, such as backlight options, can be controlled remotely.

SAP NetWeaver Mobile secures data in a number of ways. Prior to transport, data is compressed then transported utilizing secure socket layers and server certificates. Data stored on mobile devices can also be encrypted. Authentication can take place between users and mobile devices and again between mobile devices and servers. Single sign-on and centrally enforced password changes add additional levels of security [14]. A "poison pill" function is also available. If a device goes missing or is stolen, or a password is compromised, central administrators are able to destroy data on the mobile device the moment the device attempts to make network contact.

II) Components:

Android is a component driven development. For each component a custom view is created example we take a component called Horizontal Slider based on progress Bar. This slider will allow the user to move the slider back and forth on the screen and get notified when this happens. So for each visual indicator it is necessary to create an interface and view. The components include:

- Menu Builder
- Horizontal Slider
- Data Access

1. Menu Builder:

It provides a menu to the user with navigation options and enables the user to make a selection. It provides the following functionalities

- Define the business objects to be accessed by the user
- Define range of values for the business objects for the users
- Define the skins for the business objects.

2. ***Horizontal Slider:***

This component provides configuration, read, redraw etc options to the user about the visual indicator horizontal slider.

3. Data Access:

This component retrieves data from commercial databases like SAP, ORACLE APPLICATIONS etc that meet the user need. Each business object to be accessed by the business users requires to be fetched from the SAP database. This includes

1. Define the primary business objects
2. Define mapping of business objects to SAP data warehouse hypercube elements
3. Setting up rules for periodic updating of data from transaction databases into the hyper cubes and the business objects

III) Types of Users:

The system envisages two types of main users

1. Business User---depicted as user
2. Administrator---depicted as Admin

1) Business User:

The role of business user is to login to the application using Android compatible phone. The specifications of the Android phone to be used for development and testing are specified in Table1. The same programs also can be called from desktop connected to the server via internet.

2***) Administrator:***

The role of the administrator is to

1. Provide security environment for the data and users
2. Define profiles and roles of the users
3. Make available the Data from SAP system BIW (Business Information Warehouse) to the business objects on the mobile by providing the mapping functions.

VII. IMPLEMENTATION OF DECISION CONTROL PARADIGM

Decision Control Paradigm can be implemented stepwise as follows:

1. START APPLICATION

User starts the application by clicking T**he Business Console application** icon on the android mobile screen. This shall present the user with a login Screen for secure login to the application as shown in figure5.

Figure 5: Android Phone

2. LOGIN

The login screen consists of two fields

1. Name
2. Password

And a LOGIN button. User name and password as shown in figure6.Password should not be less than 8 Characters . Clicking the login button shall call the method

Authenticate_user (Username, Phone number and Password)

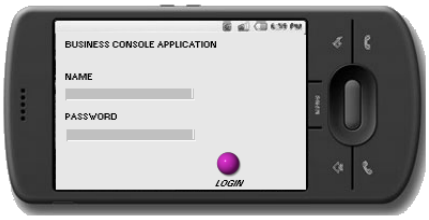


Figure6: Login Screen

This method shall read the server connection parameters as shown in figure7 from a local text file and shall call the method on the remote server for execution. The parameters shall be encrypted with phone number as key. The return value will be either 0 (authentication failure) or 1 (Authentication success). On failure of authentication, a warning message shall be displayed on login screen. (LOGIN Failed. Retry!) as shown in figure7. On successful login, Operational menu shall be displayed by calling the method

Display_menu (User)

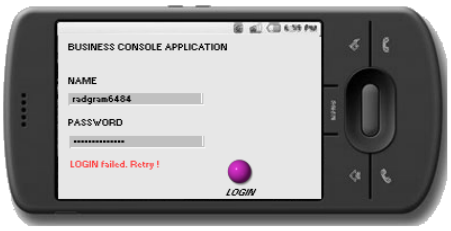


Figure 7: Login Screen with Testdata

3. OPERATE MENU

An array of objects and their business names for which the user is eligible to operate shall be fetched from database and displayed in the menu after successful user authentication. The user names of business objects shall be presented as a scrollable list on the screen with 2 buttons at the bottom – one to EXIT the program and the other to proceed to display the business Object as shown in figure8.

Figure 8: Menu Screen

4. Display Business Object

Business Objects are classified as different categories for display.

They could be displayed with different skins

1. Classical Round meter as shown in figure9:

Figure9: Classical Round Meter

2. Horizontal Slider like display as shown in figure10:

Figure 10: Horizontal Slider

3. Blinker as shown in figure11.

Figure 11: Blinker

The objects could display types

1. Instantaneous values
2. Cumulative Values
3. Levels / Range bound

Display_Object method shall be called with the following 10 parameters

1. Type of skin (CRM, THR, BLI etc.)
2. type of display (I, C, L etc.)
3. Business object ID (SAP object ID in the datawarehouse)
4. Business name (Heading to be used)
5. range ID (Scope like dept, group etc.)
6. Scan interval (in munutes)
7. user_id (Login user ID – will filer all values on this id)
8. Max Value (Max control value allowed)
9. Min Value (Min. control value)
10. Current Value
11. Auto refresh (true or false – based on which the meter dynamically refreshes.)

VIII. CONCLUSIONS

The use of smart phones is growing at an unprecedented rate and is projected to soon pass laptops as consumers' mobile platform of choice. The proliferation of these devices has created new opportunities for mobile researchers; however, when faced with hundreds of devices across nearly a dozen development platforms, selecting the ideal platform is often met with unanswered questions.

Android has a rich set of APIs, The way Android attempts to accommodate the vast expansion of sensor capabilities and user expectations of what a device should be able to do suggests a quantum jump in operating systems may be possible. The controlled isolation of individual applications should lead to much better security as well. Android powered devices that will flourish in the near future.

REFERENCES

[1] Android - An Open Handset Alliance Project, 2007. http://code.google.com/android.

[2] Earl Oliver David R. Cheriton A Survey of Platforms for Mobile Networks Research, IEEE Dec 2008.

[3] Ducrohet, Xavier (15 September 2009). "Android 1.6 SDK is here". http://android-developers.blogspot.com/2009/09/android-16-sdk-is-here.html

[4] Wayne Jansen, Tom Karygiannis, Michaela Iorga, Serban Gravila, and Vlad Korolev, Security Policy Management for Handheld Devices, The 2003 International Conference on Security and Management (SAM'03), June 2003.

[5] Gopalaswamy Ramesh; Ramesh Bhattiprolu (2006). *Software maintenance : effective practices for geographically distributed environments*. Informationweek.

[6] Write Once, Run Anywhere – A Survey of Mobile Runtime Environments S¨oren Blom, Matthias Book, Volker Gruhn, Ruslan Hrushchak, Andr´e K¨ohler Applied Telematics/e-Business Group, Dept. of Computer Science, University of Leipzig.

[7] Smart Phone Wars: Apple vs. RIM vs. ... the Android Operating System? http://blog.changewave.com/2008/04/smart_phone_apple_rim_google.html

[8] Montgomery, Justin (July 8, 2008). "T-Mobile's HTC Dream, The First Android Phone?". InformationWeek. http://www.informationweek.com/blog/main/archives/2008/07/tmobiles_htc_dr.html.

[9] "What is Android?". *Android Developers*. 21 July 2009. http://developer.android.com/guide/basics/what-is-android.html.

[10] Alex Olwal. Lightsense: enabling spatially aware handheld interaction devices. volume 0, pages 119–122, Los Alamitos, CA, USA, 2006. IEEE Computer Society.

[11] Srinivas, Davanum (2007-12-09). "Android - Invoke JNI based methods (Bridging C/C++ and Java)". http://davanum.wordpress.com/

2007/12/09/ android-invoke-jni-based-methods-bridging-cc-and-java/.

[12] R. V. Rai. Soot: A Java bytecode optimization framework. Master's thesis, McGill University, 2000.

[13] Darryl K. Taft (2005-05-20). "Eclipse: Behind the Name". *eWeek.com*. Ziff Davis Enterprise Holdings. http://www.eweek.com/c/a/Application-Development/Eclipse-Behind-the-Name.

[14] Core Security Technologies, *Multiple vulnerabilities in Google's Android SDK*, http://www.coresecurity.com/content/advisory-google

◈◈◈

Ovate Management System
With Effect of Kerberos In Organization

Javad Sharifi Boroujeni & Mehdi Hojabri

Department Of CMS Andhra University Department of CS and SE Andhra University, Vizag, India
E-mail :hami_1380@yahoo.com,hozhabri64@gmail.com

Abstract-Today, all the private and public companies in different countries to quickly preparing for major changes. Senior managers and heads of companies have special sensitivity in the structure of their company. Manager's role as leader in management and control is of considerable importance. In order to achieve successful and efficient organization where creative staff as the key and most important organizational resources are considered. Quiet, safe and without stress is required, and this gift from competent and knowledgeable managers of knowledge management is achieved. Managers are taking steps in order to achieve organizational goals, while their role in maintaining the excellence and effectiveness of organization that is emphasized by all management scholars. Where the strategic thinking of management turmoil since the emphasis on order forms. But in some organizations can be seen that not only the effectiveness and efficiency of managers do not follow, But also contradictory and destructive functional to the foundation and corner stone of all organizational structures are being sought. The rational planning organization is determined to destroy. Therefore, staffstructureandlayoutandthe on going security and stability in the organization are essential. We want introduce a new management system and improve Kerberos security method from networks to organizations.

Keywords: *organization - ovate management - eggs- systems –Kerberossecurity*.

I. INTRODUCTION

Organization: organization is the arrangement of the larger units of meaning in a paper. That's one of the things that are going to be very different from one course to the next. Organization typically refers to the Large Elements of text structure. Sometimes these elements are formalized in practice, as in the typical lab report, through consistent use of headings. Sometimes elements of organization are only informally acknowledged -like the thesis of an academic paper. An organization is a social group which distributes tasks for a collective goal. The word is derived from the Greek word *organ on*, itself derived from the better-known word *argon* which means "organ" - a compartment for a particular task. There are a variety of legal types of organizations, including corporations,governments , non-governmental organizations, international organizations, armed forces, charities, not-for-profit corporations, partnerships, cooperatives, and universities.

Kerberos: Kerberos is a secure method for authenticating a request for a service in a computer network. Kerberos was developed in the Athena Project at the Massachusetts Institute of Technology MIT. The name is taken from Greek mythology; Kerberos was a three-headed dog who guarded the gates of Hades. Kerberos lets a user request an encrypted "ticket" from an authentication process that can then be used to request a particular service from a server. The user's password does not have to pass through the network. A version of Kerberos client and server can be downloaded from MIT or you can buy a commercial version.

*System:*A system is a set of interacting or interdependent components forming an integrated whole. A system is a set of elements and relationships which are different from relationships of the set or its elements to other elements or sets. Fields that study the general properties of systems include systems theory, cybernetics, dynamical systems, thermodynamics and complex systems. They investigate the abstract properties of systems' matter and organization, looking for concepts and principles that are independent of domain, substance, type, or temporal scale. Most systems share common characteristics, including:

- Systems have structure, defined by components/elements and their composition;

- Systems have behavior, which involves inputs, processing and outputs of material, energy, information, or data;
- Systems have interconnectivity: the various parts of a system have functional as well as structural relationships to each other.
- Systems may have some functions or groups

A set of detailedmethods, procedures, and routines established or formulated to carry out a specific activity, perform a duty, or solve a problem. An organized, purposeful structure regarded as a whole and consisting of interrelated and interdependent elements. These elements continually influence one another directly or indirectly to maintain their activity and the existence of the system, in order to achieve the goal of the system.

II. INSTRUCTION

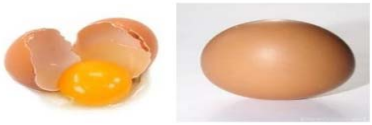
Define a management system based on physical characteristic so f the eggs that is called ovate management system. It must first know the physical properties of eggs in order to implement our new organizational structure.

Protectivemechanisms of egg: As the fertilized egg in the reproductive cells, various materials have been prepared from a chicken that is achieved, it would be astonishing if it is to protect the amenities of the influx of various factors including microbes normally be supplied. Physical and chemical defenses of eggs placed in two sections.

- Physical defenses of eggs, shell, shell membranes, bag albumin
- Chemical defenses of eggs, shell membranes, albumen

Defensiveshieldshell: Physiological basis of the building shell includes several compound sthatasa barrier to gas exchange andwaterisvital. An internal thin stripFilm, which preventsthe shell, is the innermostlayer of the protein by internal and external shell membranes is covered. Thin inner layer and the inner shell membrane, preventing there lease of oxygen are together, but the shell, in general inhibits is the releaseof carbon dioxide and water vapor. Limestone core bumps formed on the crystals grown from a small upward. Connected to adjacent crystals and crystalline lime stone columns make up the conecells of the sponge layer or layers cover afence. Healthy skin pores and cones where the edges are not completely together are formed. Therefore, the mechanical strength and quality of their cones and releaseMamylary breathable shellis concerned. Covering shell cuticle is uneven. Shell egg protection the embryo inside the egg plays an important role. Site frespiratory gas exchange and maintain moisture in eggs and embryos as amechanical shield acts as a scaffold. However, the absolute shell of a protective barrier gainst bacterial penetration is not. The only available route for germ stop enetrate into the pores of the shell is in tact. Events include events hokey flattened form is non-split unitsmake up the channels and penetrate into the crystal layer, and furcatedendsareadjacentto Calcareousbumps. Cuticleon the shell material at the time of spawning has not specified the construction and building of the vesiclesare formedwithin minutes ofthe firstonsetof spawning.

Defensiveshieldshellmembranes: Shell membranes under the shell matrix, there are three separate layers. Inner and outer membranes that contain a network of fibers that have a random orientation and the third layer as a uniform layer, and Electron-dense material containing buffer or barrier membranes Limited membrane called. Membranes in gas exchange and movement of calcium the shells are involved. in the osmotic pressure of the shell membranes, flooding is possible in the pores and move it along the channel, continuous flow displacement caused by bacteria which occur during the egg coating.The mostimportantline of defense againstbacterialmembranesistransportandactasafilteragainstmicrobes. Themembraneof bacteria andfungithathavepassed throughthecuticleandshell, to thetraps. Growthofthe fetusinside theeggwithout anykindof pollutioninitsinternalcontents, despitethe presenceofbacteriaontheshellmembranesis possible.



III. ORGANIZATIONALDESIGN

Concepts, ToolsandModelsOrganizationstructureasamanagement toolforimplementing thestrategies andgoals.Whyisorganization structuretoolamongmanagers.

- Itisrelatively easyto change behavior change, modify, or cultureis more difficultto implementthe newstrategy. Butitsattractivenesstoitssevereimpactonthereturns.
- Structurecan beregardedasa major concern ofthe newmanagement.
- Withthechange, the newmanagercan beleftin the organizationaffectstheirpersonaleffects.

- Changing thestructureorstyle ofmusiccould benewform oforganizationwasoperating.
- Withstructural changes, newchangescanbecommunicatedinthe organization.

In generalwe can say:

A. Thedesignispart of thejobmanager.

B. Shapingthe design ofthestructure, configurationandmorphologyare concentrated on theorganization structure.

C. Tobe effectivemust be betweenthetheoreticalaspectswithpracticalaspects ofa balanceis established.

D. Managersshouldlookcontinuallybetween thetwostructures.

- Strategicperspective
- Look attheirsocial and culturalbalance.

Designermustconsidertwoquestions:

A. Howtostructurethe implementation ofvariousstrategiesandmanagementenablesorganizations to performtherequired workwill befacilitated?

B. How the people affects withinthe organization andpatterns ofinformalorganizationstructureaffectsbehavior.

IV. TYPES OFORGANIZATIONALDESIGN

A. The organization's strategicplan

- Composition of organizational units
- Reporting relationships
- Other structural links between units
- Information systems, organization and control measures
- Macroorganizational procedures and practices
- Techniques of working

B. PlanofOperations

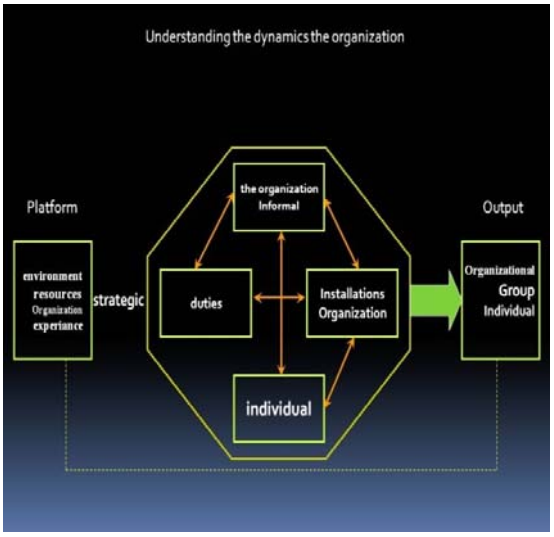
- Sub units of work resources tools, materials
- Reward systems of working units
- Physical environment in the sub units
- Design jobs

Sourceofcriteriadeveloped:

- Strategy
- Responsibilities
- Pathologicalsituation
- Otherinformation

Thedecisionontheplan:

- Thedesigncriteriadeveloped
- Developeda variety ofstrategicgroupings
- Evaluation ofstrategicgroupings
- Coordinationrequirements ofevaluation
- Formingthe structuralmechanisms
- Mechanisms to evaluate the structural link
- Perform effectiveness analysis
- Refining and defining the project
- Identify operational issues facing plans
- Performed to identify the issues Facing



V. THE CONCEPTOFORGANIZATIONAL STRUCTURE

Organizational structure, systemsthinkingismanifested. Organizationcomposed ofelements, relations between theelementsandstructure of relationsasawholethatmakes upaunit. The bangSaied structure, thesupremecombinationofthe relationships betweenorganizationalelementsthatwillshape thephilosophyof organizationalactivities.Systematthe organizationstructureshows thatthe compositestructureof thehardwareelementson the one handandtheotherissoft. Afterthehard, tangibleelementssuch asgroups andorganizational unitsarehierarchical. Relationships betweenthese unitsand groupsinorganization

structureareманifested inyourelement.Existing literature from different angles to look at the structural relationships. Were identified in three dimensions: the hierarchical, the special work and then a center of learning and study is exceptionally unique. He studied in three dimensions are as follows:

Following hierarchy: the relative ranking organizational units similar to the way an organization chart shows.After a special task: different kinds of things that should be done in the organization. *After inclusion and centrality:* how far or near the core of every individual in the organizationrepresents the organization. The correct combination of the only formal structure that is mainly manifested in the organizational chart. However, the reality is that there are a range of organizational forms that they cannot be explained simply through the organizational chart.

VI. THE STRUCTURAL EVOLUTION

The centralissueis whetherthestructuralstudiesfollowthe structureofcorporatestrategyleadstobetter performanceor not?Contingencyrelationshipbetween theenvironment, the specialworkandtheorganization,theclassicstudies ofthestructure-theperformance. Studieson theconditionsnecessary forformingtheexternalvariablehavethe appropriatestructuralforms.Layers ofthehierarchy: the organizationalperspective, bythe highestauthorities will benotified. Specialfocusonorganizational unitstodeterminethe exactmannerand ISobserve thework ofspecialization. Strongconcentration of powerandmanagementcontrol: theverticalrelationshipbetween theagencyheadsto coordinateactivitiesandorganizationalbase ofthe pyramidis used. Highlevels offormalization: There areseveraldrybureaucraticrulesandproceduresgoverning thepracticeandthere islittleindividual freedom. Showsthe evolution oforganizationalstructure, hierarchical structuretoa moreflatstructureandismore flexibleandmodernbusinessworlditisreplaced byanorganicstructure.

*Organicevolution:*Organicstructureprovidesasimulation modein whichcomplexorganizationsand socialphenomenaareperceived.

Organicstructurefeaturesinclude:

- FlatandteamorientedIn Part.
- Decentralizationof powerand control.
- High levelofformalization.

*Competitive advantageinknowledge-based economy:*Parallel to theentryof companiesintothe 1990s, studentsbecameone of themost importantstrategicresources. Generateknowledgeforsustainablecompetitive advantageandorganizationalsuccessisessentialaxial.

VII.EFFECTIVE KNOWLEDGE MANAGEMENT REQUIRESTHE FOLLOWING

*TheBorder***:**firmsbasedorganizationsshouldbefreeofp hysicalboundariesandlimitations ofaconceptual frameworkbased onorganizationalidentity andtrustarecreated. Informalrelationsplay an important roleindiminishing the coloroftheborder. More fluid, effectiveknowledgemanagementrequiresknowledge oftheplaceisa repositoryof knowledge.

*Attractions:*Effectiveknowledgemanagementis basedlargely onimplicitknowledgemanagement. Informalrelationships, the interaction betweentaskandinter-organizational sharing oftacit knowledgeas well asupgradedandthe mainwaytacit knowledgeofmoldingandproduction.

Flexibility: To generateeffectiveknowledgebasedartifacts, the structuremustbeflexible.

VIII.SPECIFICATIONMANAGEMENT OVATE ORGANIZATION

Due tothe uniquephysical propertiesofshelleggs,webuiltan organizationbased onthese characteristics;itmeansare intrinsically ovate organization. The first point intheegg shellithasa uniquecrystal structurethatwe areexactlyontheoutershellmadeof theorganization.Organizationto maintaintheirsurvivalin addition tohaving adefinedrelationshipbetweenorganizational membersneedstobeconcerned withthe environmentoutside the organization. Individualororganizationrelationshipsexternal environmentthatarewithin theorganization, orsupportersofthe organizationare classified ascompetitorsorenemies. Theinnershell ofdutyoutside theorganizationto identifypeoplewhoaregoingtobeorganized. Andissuanceofpermitsfortheirentryintothe organization. Number ofentry and exit, and the permeabilityofthe organizationtheywill bedetectedbythe innerlayer. Theoutershellofthesecurity forcesorhasconservationorganization. Theinnershellofmiddle managersandmiddlemembers ofthe organizationarein place. Shellin order to havestrongorganizationalcontrol oftheofficeisresponsible forallsecurity forces. Our means

of internal crust of organization is middle members and middle managers. This is second important difference among this organization with regular organization. In this kind of organization we change position of down operation level with middle operation level. Intermediate level of work has alessexten to the organization but they have key tasks, themiddlelevels of the organization came around of Protectivecrust to increase security, And down level operation with wide work operation come among middle level and chief area , in this organization amount of responsibility of Presidential field and middle level will be increase. In this kind of organization ovate organization Surface pressure tolerance will be increase also till that the organization is resistant against all vertical pressure that we can making categorize this pressure like:

- Lowered profit organization
- Sanctions or conditions governing from out of organization.
- The sudden change in market share
- Out numbering widespread of Competitors in the market.
- Intent to destroy the organization.
- The sudden increase in costs.
- The sudden change in policy.

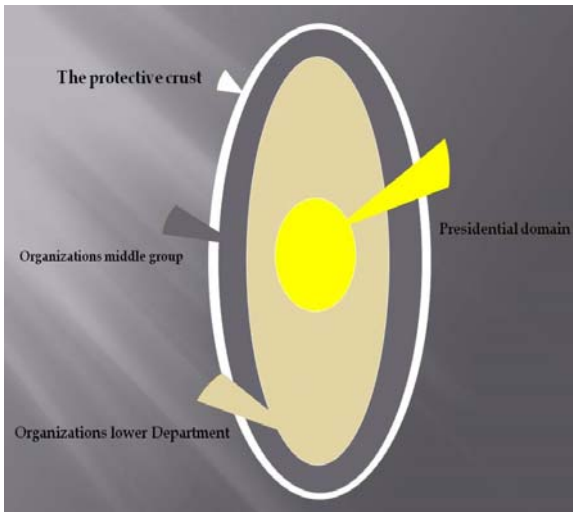
IX. IMPROVEMENT OF THE OVATE ORGANIZATION

All the ovate organization is exposed to various horizontal pressures.

All the ovate organizations in order to the increase of their tolerance in horizon level need to regard some important matter:

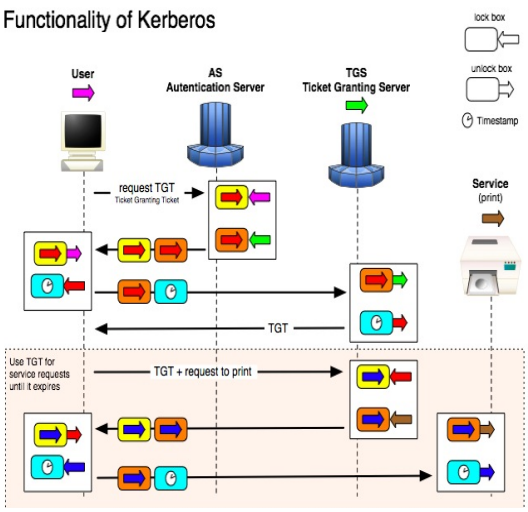
- Make coordination among work groups in organization.
- Make satisfaction in organization.
- A relationship between the heads of organizational cross and employee.
- Inter-organization coordination without side organizations.
- Considering the particular organization for entry into organization in crisis.

In this organization middle manager must do more activity rather than other organization in this kind of organization we can have a gold ring that all of members can be formed from down managers, middle managers and Presidentialdomain, and also we have a platinum ring in order remove high level problem in organization that can be formed from Presidentialfield and middle managers. In ovate organization all high level position can be accept by dean of every section.



X. BRING AFFECT OF KERBEROS MODEL IN OVATE ORGANIZATION

Kerberos model is one model from it in computer science weuse of that in order to increasing security in inner and extern environment of organization. In Kerberos system everybody that is available in organization must defined in system and according their position they will get a card or ticket and their card or ticket will be credit according to their job and their position, hence they have access to information and facilities is available in organization. In this state we can be sure nobody can not access to all information of organization in order abuse of them. In Kerberos model we try also to knowing external people that have continuous relation with organization and issue a ticket for them. This process can be help us to make satisfied our customer and made sure us they cannot come inside of our organization more than their allowance. Kerberos model can be developing to all organization in order to increase of security and accessibility. In this model also we can defined that our staff or costumer can be accessibility to information from specific connection or specific system also, in this case nobody cannot access to information with their card or ticket from out of defied condition.



REFERENCE

[1] ACKEOFF, R.L (1994)," The Democratic corporation", Oxford University Press, London.

[2] AHMED, P.K (1998)."Culture and Climate for Innovation". European Journal of Innovation Management, Vol1, No1, pp30-43.

[3] BIELY,P.E,KESSLER,EH and CHRISTENSEN,E.W(2000)"Organizational Learning. Knowledge And wisdom, journal of Organizational Change"

[4] BUNGE, M. (1976), "A World ofSystems, Reidel, Dordrecht.

[5] JONATHAN OLIVER' The Call of Kerberos' – (Abaddon Books)

[6] MIKAEAL SVEREON, Working with Microsoft® FAST™ Search Server 2010 for SharePoint

[7] Peebles, E. D. and McDaniel C. D. (2004). A Practical Manual for Understanding the Shell Structure of Broiler Hatching Eggs and Measurements of Their Quality. Department of Poultry Science Mississippi State University.

[8] Board, R. G. and Tranter, H. S. (2002). The Microbiology of Eggs. In: Egg science and technology. eds. Stadelman, W. J. and Cotterill O.

[9] Kerberos" A Network Authentication System", Brian Tung

◈◈◈

Unicode Optical Character Recognition and Translation Using Artificial Neural Network

Pramod J Simha & Suraj K V
Electronics and Communications dept., BNM Institute of Technology, affiliated to VTU, Bengaluru, India
Email: jsimha54@gmail.com & surajkv@in.com

Abstract **-** In this paper we demonstrate the capabilities of Artificial Neural Network implementations in recognizing extended sets of optical language symbols. We describe an advanced system of classification using probabilistic neural networks. Training of the classifier starts with the use of distortion modeled characters from fonts. Statistical measures are taken on a set of features computed from the distorted character. The space of feature vectors is transformed to the optimal discriminant space for a nearest neighbor classifier based on these classifiers. In the discriminant space, a probabilistic neural network classifier is trained. For classification, we present some modifications to the standard approach implied by the probabilistic neural network structure which yields significant speed improvements. An emerging technique in this particular application area is the use of Artificial Neural Network implementations with networks employing specific learning rules to update the links (weights) between their nodes. One such network with supervised learning rule is the Multi-Layer Perceptron (MLP) [1] model. It uses the Generalized Delta Learning Rule for adjusting its weights and can be trained for a set of input and desired output values in a number of iterations. The project has employed the MLP technique mentioned and excellent results were obtained for a number of widely used font types. The approach involves processing input images, detecting graphic symbols, analyzing and mapping the symbols and training the network for a set of desired Unicode characters corresponding to the input images.

I. INTRODUCTION

The classic difficulty of being able to correctly recognize even typed optical language symbols is the complex irregularity among pictorial representations of the same character due to variations in fonts, styles and size. This irregularity undoubtedly widens when one deals with handwritten characters.

Modeling systems and functions using neural network mechanisms is a relatively new and developing science in emerging technologies. The particular area derives its basis from the way neurons interact and function in the natural animal brain, especially humans. The animal brain is known to operate in massively parallel manner in recognition, reasoning, reaction and damage recovery. The typical human brain at birth is estimated to house one hundred billion plus neurons. Such a combination would yield a synaptic connection of 1015, which gives the brain its power in complex spatio-graphical computation [1] [2]. Neural networks have seen an explosion of interest over the last few years, and are being successfully applied across an extraordinary range of problem domains, in areas as diverse as finance, medicine, engineering, geology and physics. Indeed, anywhere that there are problems of prediction, classification or control, neural networks are being introduced.

Hence the conventional programming methods of mapping symbol images into matrices, analyzing pixel and/or vector data and trying to decide which symbol corresponds to which character would yield little or no realistic results. Clearly the needed methodology will be one that can detect proximity of graphic representations to known symbols and make decisions based on this proximity. To implement such proximity algorithms in the conventional programming one needs to write endless code, one for each type of possible irregularity or deviation from the assumed output either in terms of pixel or vector parameters, clearly not a realistic fare.

II. MULTI LAYER PERCEPTION (MLP)

The basic block of a neural network is an artificial neuron whose function can be defined by a two-step procedure.

First it receives a number of inputs (either from original data, or from the output of other neurons in the neural network) [3]. Each input comes via a connection that has strength (weight). Each neuron also has a single threshold value. The weighted sum of the inputs is

formed, and the threshold subtracted, to compose the activation of the neuron (also known as the post-synaptic potential, or PSP, of the neuron). Next, the activation signal is passed through an activation function (also known as a transfer function) to produce the output of the neuron.

This describes an individual neuron. If a network is to be formed, there must be inputs (which carry the values of variables of interest in the outside world) and outputs (which form predictions, or control signals). However, there also can be hidden neurons that play an internal role in the network. The input, hidden and output neurons need to be connected together. This forms a feed forward network as in *Fig 1*.

The Multi-Layer Perceptron Neural Network is perhaps the most popular network architecture in use today. The units each perform a biased weighted sum of their inputs and pass this activation level through an activation function to produce their output, and the units are arranged in a layered feed-forward topology. The network thus has a simple interpretation as a form of input-output model, with the weights and thresholds (biases) the free parameters of the model. Important issues in Multilayer Perception (MLP) design include specification of the number of hidden layers and the number of units in each layer.

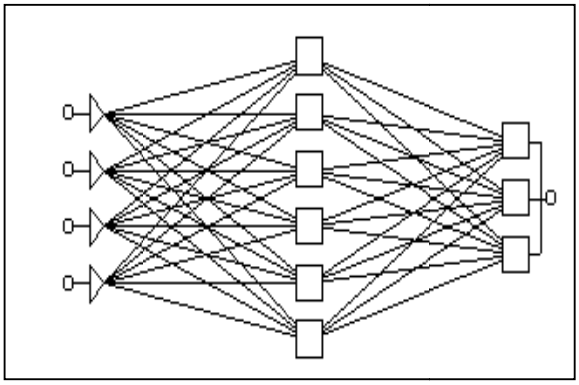


Fig. 1 Typical feed forward network

Here we use a 3 layer network consisting of an input layer, a hidden layer and an output layer. The input layer is not really neural at all: these units simply serve to introduce the values of the input variables. The hidden and output layer neurons are each connected to all of the units in the preceding layer. Again, it is possible to define networks that are partially-connected to only some units in the preceding layer; however, for our applications fully-connected networks are better. Most common activation functions are the logistic and hyperbolic tangent sigmoid functions. The project used the hyperbolic tangent function:

$$f(x) = \frac{2}{(1+e^{-2x})} - 1$$

and derivative:

$$f'(x) = f(x)[1 - f(x)]$$

III. NETWORK FORMATION

The MLP Network implemented for the purpose of this project is composed of 3 layers, one input, one hidden and one output as in *Fig2*. The input layer constitutes of 150 neurons which receive pixel binary data from a 10x15 symbol pixel matrix. The size of this matrix was decided taking into consideration the average height and width of character image that can be mapped without introducing any significant pixel noise. The hidden layer constitutes of 250 neurons whose number is decided on the basis of optimal results on a trial and error basis. The output layer is composed of 16 neurons corresponding to the 16-bits of Unicode encoding. To initialize the weights a random function was used to assign an initial random number which lies between two preset integers. The weight bias is selected from trial and error observation to correspond to average weights for quick convergence. This is done after initial trails with unity weights are done.

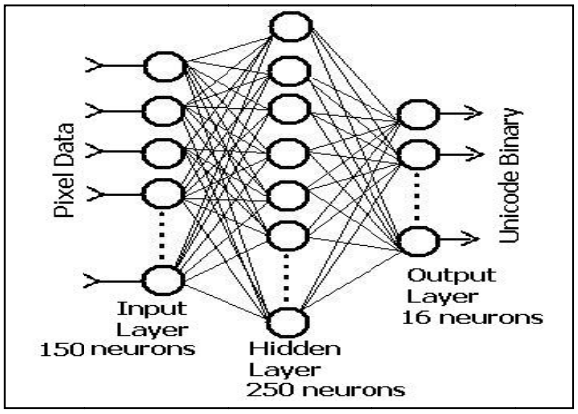


Fig. 2 Network Formation of MLP network

IV. SYMBOL IMAGE DETECTION

The process of image analysis to detect character symbols by examining pixels is the core part of input set preparation in both the training and testing phase. Symbolic extents are recognized out of an input image file based on the color value of individual pixels, which for the limits of this project is assumed to be either black RGB (255,0,0,0) or white RGB(255,255,255,255). The input images are assumed to be in bitmap form of any resolution which can be mapped to an internal bitmap object in the Microsoft Visual Studio environment. The

detection involves algorithms for Determining character lines, Detecting Individual symbols and Symbol Image Matrix Mapping [4]. Since we use a single dimensional network an additional algorithm for converting two dimensional matrixes to a single dimension matrix needs to be implemented as in *Fig 3a*. The algorithm individually yields the following results as represented in the figures.

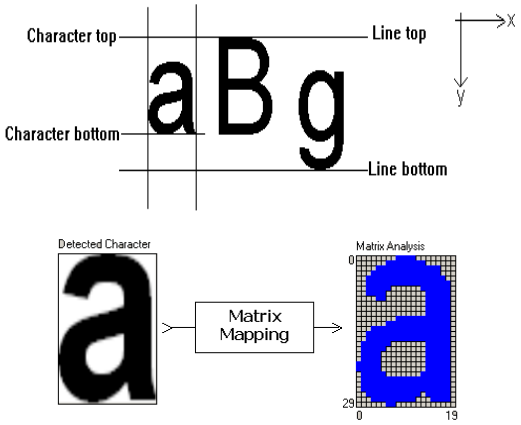


Fig. 3a Line and Character boundary detection

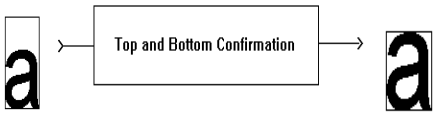


Fig. 3b Top and bottom confirmation

Mapping symbol images onto a binary matrix. The procedure also assumes the input image is composed of only characters and any other type of bounding object like a border line is not taken into consideration as in *Fig 3b*.

V. TRAINING

Once the network has been initialized and the training input space prepared the network is ready to be trained. Some issues that need to be addressed upon training the network. This includes the extent of chaotic nature of the input space, the complexity of the patterns which need to be recognized which is measured in terms of feature overlap and data size, the number of Iterations (Epochs) needed to train the network for a given number of input sets and error threshold value must be used to compare against in order to prematurely stop iterations if the need arises. All these issues can be addressed by setting values for Learning rate, Sigmoid slope [5] [6], number of epochs and Weight bias. The Flow chart is as shown in *Fig 4*.

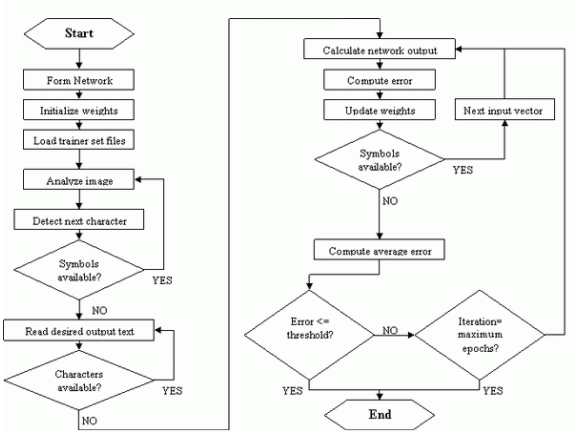


Fig. 4 Flow chart of training the network

For the purpose of this project the parameters uses are: Learning rate = 150, Sigmoid Slope = 0.014, Weight bias = 30 (determined by trial and error), Number of Epochs = 300-600 (depending on the complexity of the font types), Mean error threshold value = 0.0002 (determined by trial and error)

VI. TESTING

The testing phase of the implementation is simple and straightforward. Since the program is coded into modular parts the same routines that were used to load, analyze and compute network parameters of input vectors in the training phase can be reused in the testing phase as well. This has also been worked upon in MATLAB [7] and the flow chart is as shown in Fig 5.

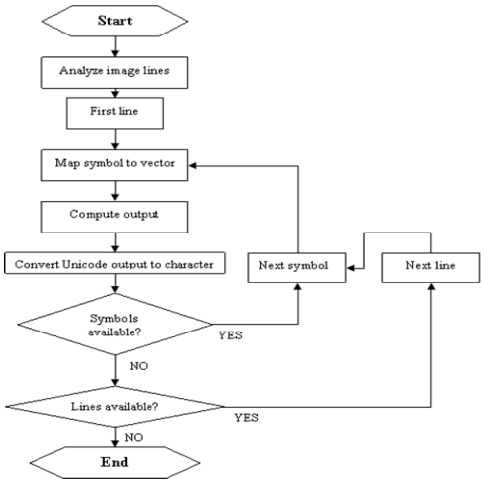


Fig. 5 Testing input images for characters

VII. TRANSLATION

This module is built using Google API for .NET which basically translates the output of Unicode to required foreign language. This uses the already method available in dynamic link library called as Translate Client. The client calls the server for translation and retrieves back the requested foreign language.

VIII. IMPLEMENTATION

The software is built using Microsoft's Visual C# .NET 2008 Express Edition. All the modules are programmed to optimum level. Also testing, training and translating are verified. A screen shot of the software is shown in Fig 6.

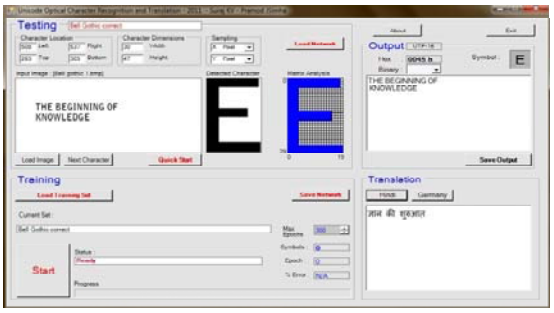


Fig. 6 Screen shot of software implemented for OCR

IX. RESULTS

The network has been trained and tested for a number of widely used font type in the Latin alphabet. To study the variation due to each parameter, that particulate parameter is varied keeping all other's constant.

A. Result for variation in number of Epochs (Iterations)

Number of characters=90, Learning rate=150, Sigmoid slope=0.014

Font Type	**300**		**800**	
	№ of wrong characters	% Error	№ of wrong characters	% Error
Latin Arial	4	4.4	1	1.1
Latin Tahoma	1	1.1	0	0

B. Result for variation in number of Input characters

Number of Epochs=100, Learning rate=150, Sigmoid slope=0.014

Font Type	**20**		**90**	
	№ of wrong characters	% Error	№ of wrong characters	% Error
Latin Arial	0	0	11	12.2
Latin Tahoma	0	0	8	8.89

C. Result for variation in learning rate parameter

Number of characters=90, Number of Epochs=600, Sigmoid slope=0.014

Font Type	**50**		**120**	
	№ of wrong characters	% Error	№ of wrong characters	% Error
Latin Arial	82	91.1	3	3.3
Latin Tahoma	56	62.2	1	1.1

Effect of changing various parameters is studied and the influence of each variation is as given below.

A. Increasing the number of epochs has positive proportionality relation to the performance of the network. However in further increasing the number of epochs has an adverse effect of introducing more wrong recognitions. This is called overlearning.

B. Increase in number of Input characters usually has adverse effect on accuracy which is quite obvious. As the network has to compare input symbol with more symbols, the decision will be subjected to more ambiguity.

C. Size of the input symbol is also found to impact the detection process. This is due to the fact that large size of symbol is scaled to only 150 pixels and this might not be sufficient to represent all symbols distinctly.

X. CONCLUSION

Optical Character Recognition is the process of taking images of text or handwriting and converting it to digital form. This is done by a "training" and "Recognition" process in software. If there are any discrepancies, it will be corrected and the software will save that character for future matching. OCR is a very helpful tool, with lots of implementation possibilities. What you otherwise take a lot of time to do manually, is a relatively quick process. OCR allows for any user to convert a document accurately.

The basic idea of using extracted features to train an ANN seems to work although the success rate is not impressive. There are several possible changes that the current bottleneck for speed is the feature extraction stage. The other obvious step is to increase the training data set. This requires some effort, but clearly more training data will lead to a more robust and accurate ANN. Some other fairly trivial features are still missing. For example, characters like the apostrophe (') and comma (,) look very similar and can be distinguished only by vertical location.

The most used applications of OCR are from Adobe Acrobat and Microsoft OneNote. Google also uses reCAPTCHA to digitize books and printed materials. Instead of hiring people to check words that are unknown, it incorporates a "Bot" prevention program. In order to access material in a site, or sign up for a service, users must input words into a text field.

APPENDIX

Our work in the project proceeded in group meetings with contribution and inputs from each person in each of the sections of the project. Working as a group encouraged great ideas through open discussions.

REFERENCE

[1] *Using Neural Networks to Create an Adaptive Character Recognition System,* 2002, Alexander J. Faaborg Cornell University, Ithaca NY

[2] S. Haykin, *Neural Networks: A Comprehensive Foundation,* Macmillan College Publishing Company, New York, 1994.

[3] *Neural Networks and Fuzzy Logic*. 1995, Rao, V., Rao, H. MIS Press, New York

[4] *Artificial Intelligence and cognitive science*. 2006, Nils J. Nilsson, Stanford AI Lab http://ai.stanford.edu/~nilsson

[5] *Hand-Printed Character Recognizer using Neural Network*. 2000, Shahzad Malik

[6] Neural Networks Algorithms Applications and Programming Techniques. 1990 Christof Koch, *California Institute of Technology*

[7] Hahn, B. (2002). *Essential MATLAB for scientists and engineers.* University of Cape Town

◈◈◈

An Application of Neural Network And Genetic Algorithm : Evidence From India

Tuhin Mukherjee[#1] & Monosri Banerjee[#2]

[# 1] Department of Business Administration , University of Kalyani, Nadia, West Bengal, India
[# 2] Heritage Institute of Technology, Kolkata, West Bengal, India
E-mail : [1]tu_2002@rediffmail.com & [2] monosri2807mou@gmail.com

Abstract - This paper is an attempt towards financial prediction in Indian stock market over recent years, using Artificial Neural Network (ANN) and Genetic Algorithm (GA). This model is named Genetically optimized Neural Network (GNN). We have tested this newly created model against traditional ARCH/GARCH models using z-test. We have used different error metrics like Average Absolute Error (AAE), Mean Square Error(MSE), Max AE, during our comparative study. This paper concludes the difference of predictive ability of our model with that of traditional ARCH/GARCH models.

Keywords-*Artificial Neural Network (ANN) forecasting models, GeneticAlgorithm(GA).*

I. INTRODUCTION

The prediction of the price movement of the stocks is one of the issues on research of stock market. Because the neural network can find a model proved by the training set, so the training set and the settings of the parameters become extremely important. In addition, because the stock market's dynamics are very quick and the model for this system may change in the short term, the more recent data should be given much weight to on the consideration of the market. On the other side, the old data should be lower estimated by the network, without loosing much of the general characteristics of the model of the domain.

While numerous scientific attempts have been made, no method has been discovered to accurately predict stock price movement. The difficulty lies in the complexity of modeling human behavior. This paper is an attempt to use ANN approach for financial forecasting (direction of price movement as well as future estimate).

The rest of this paper describes difficulties in financial forecasting and suitability of ANN modeling in this respect, review of literature, experiment of this study comprising of searching for optimal parameter selection of ANN approach, and a final conclusion indicating major findings with new area of investigation as well as limitation of this research study.

II. DIFFICULTIES in FINANCIAL FORECASTING and SUITABILITY of ANN MODELING

Stock market has long been considered a high return investment field. Due to the fact that stock markets are affected by many highly interrelated economical, political and even psychological factors that interact with each other in a very complex fashion, it is very difficult to forecast the movement in stock market.

Predicting is telling about the future which will incur certain error. To produce a meaningful prediction, the error incurred must be minimum. There are several ways used by investors to predict stock market returns such a technical analysis, fundamental analysis and mathematical models. However these techniques incapable of determining the exact forecast price. Due to these imperfection factor current studies using soft computing techniques (Soft Computing represents that area of Computing adapted from the physical sciences.) such as Granular Computing, Rough sets, Neural Networks, Fuzzy sets and Genetic Algorithms are highly used to improve the prediction accuracy and computational efficiency compared to earlier techniques.

With the advancement being made in computer and telecommunication technologies today, the world's major economies and financial markets and becoming more and more globalize. As this trend accelerates, financial markets are becoming more and more

interrelated and fundamental factors will become increasingly critical to financial market analysis. In the global marketplace, the prevailing methods of technical analysis where a single market is modeled through historical simulation and back testing of its own past price (or volume) behavior is rapidly losing its competitive advantages. Institution and individual traders both are increasingly applying new technologies to financial forecasting. Recent research shows that these nonlinear domains can be modeled more accurately with these technologies (like ANN) than with the linear statistical and single-market methods that have been the mainstay of technical analysis throughout the past decade.

Another advantage of ANN implementation is that the processing is distributed among many nodes. Even if some of the nodes fail to function properly, the effect on the overall performance of the system will not be significant. This assertion can be verified by turning off m randomly selected hidden layer nodes and observing the resulting effect on the system performance.

However due to their large number of inputs, network pruning is important to remove redundant input nodes and speed up training and recall. Essential features of a neural network are: The network topology, Computational functions, and Training algorithm. Decisions on the target output with respect to concerned inputs will select these features along with their respective parameters like learning rate, number of hidden layers, and number of nodes in each layer etc.

Financial neural network must be trained to learn the data and generalize, while being prevented from overtraining and memorizing the data. Once trained, the network parameters (weights) will be kept fixed. The model is then used with the input data set for prediction. A neural network can be designed to predict the direction, magnitude or just turning points in the stock price movement.

III. LITERATURE REVIEW

Research reviewed in this area generally attempts to predict the future data points of some time series using historical data sets. Possible time series include: Base time series data (e.g. closing prices) or time series derived from base data (e.g. indicators which are frequently used in Technical Analysis). There are many studies that attempted to predict future values of a series from the past values of that same series or using data from different series. The studies those are representative of the current research in the time series prediction include (Chan and Foo, 1995; Quah and Srinivasan, 2000; Yao and Poh, 1995; Hobbs and Bourbakis, 1995;Austin Looney *et al.* , 1997; FalasCharitou*et al.* ,1994). These studies consider data from both fundamental and technical analysis. For example Falas*et al. (1994)* used ANNs to attempt to predict future earnings based on reported accounting variables. They found no significant benefit using ANNs and concluded that accounting variables chosen were not appropriate earning predictors. Quah and Srinivasan (2000) used mainly accounting variables to predict excess returns (with limited success). Chan and Foo (1995) used ANNs to predict future time series values of stock prices, and used these 'future' values to compute a variety of Technical Indicators. They concluded that the networks ability to predict, allows a trader to enter a trade a day or two before it is signaled by regular technical indicators, and this accounts for the substantially increased profit potential of the market participants.

IV. EXPERIMENT OF THIS STUDY

The data used in this research is the daily stock prices . In our previous works, we have used closing stock prices for similar research but to check consistency of the result, the present paper uses daily opening prices . The data set covers the period from 1st Jan 2009 to 31st Dec 2011 (3 calendar years). A randomly selected 100 stocks (whichhave been participating in BSE 500 within the period of our research study) have been considered in the sample. Industry wise classification of our sample is given in table 1.0.

Main aim of this paper is to compare the predictive ability of GNN based financial forecasting models with that of ARCH/GARCH models in the context of Indian stock market.

Firstly, it is necessary to formulate null hypothesis which will be tested against its alternate hypothesis. Taking the null hypothesis that there is no difference in predictive ability of two models (i.e. GNN and ARCH/GARCH), we can formulate hypothesis as:

H_O: $\mu(GNN) = \mu(ARCH/GARCH)$

H_A: $\mu(GNN) \neq \mu(ARCH/GARCH)$

As the sample size is large, so z-test will be suitable for difference in means, assuming the populations to be normal and shall work out the test statistic z as under (C.R.Kothari,2003):

$$Z = (\mu(s_1) - \mu(s_2)) / (\sigma(s_1)^2/n_1 + \sigma(s_2)^2/n_2)^{1/2}$$

Since the population variances are not known, so we have used the sample variances, considering the sample variances as the estimates of the population variances.

Using the statistical software package SPSS (version 10.0),we get the computed z-values for different error metrics in following table 1.1.

As alternate hypothesis is two sided, so we shall apply a two-tailed test for determining the rejection regions. At 5% level of significance, it comes to as under (using normal curve area table) : R: $|z| > 1.96$.

V. RESULT OF THIS EXPERIMENT

Due to space constraint, we included only first 10 companies in table 1.1, though summary of all 100 companies is given in table 1.2. Evidently it is clear that all the observed values of z w.r.t. different error terms of each company came consistent (i.e. either all rejects null hypothesis or all fail to reject), which is quite natural. Moreover, it is observed from z statistic that 60% of the sample companies lie in the rejection region and hence reject the null hypothesis. So we can conclude that 60% of our sample stocks imply that predictive abilities of two models (GNN and ARCH/GARCH) differ significantly. This summary is represented in following table 1.2. This finding is very closed and consistent with Austin (1997).

ACKNOWLEDGEMENT

This paper is continuation of our previous research papers and really an advancement in interdisciplinary research field of computer science and financial management. All the authors contribute in this paper truly and sincerely. Valuable references are given along with this paper.

REFERENCES

[1] Abdullah, M. H. L. b. and V. Ganapathy (2002). "Neural Network Ensemble for Financial Trend Prediction". *Tencon 2000: Proceedings: Theme: Intelligent Systems and Technologies for the new millennium.*

[2] Austin, M. , C. Looney et al. (1997). "Security Market Timing Using Neural Network Models". *New Review of Applied Expert System*, Vol. 3, pp. 3-14.

[3] Baba, N. and H. Handa (1995) "Utilization of Neural Network for Constructing a User Friendly Decision Support System to Deal Stock". *IEEE* International Conference on Neural Networks

◈◈◈

Detection And Isolation of Reluctant Nodes In MANET

***M. Swetha Reddy, **Jayashree S Patil & #K. Bharath Kumar**

*Ravindra College, Kurnool, **GNITS, Hyderabad
Vignan College, Hyderabad,
E-mail : ecm.061925@gmail.com, jshivshetty@gmail.com & bls26092011@gmail.com

Abstract - An ad hoc network is a collection of various wireless mobile nodes dynamically forming a temporary network without the use of any existing network infrastructure or centralized administration. The transmission range of nodes is limited; hence nodes mutually cooperate with its neighboring nodes in order to extend the overall communication. However, along with the cooperative nodes, there may be some reluctant nodes like selfish nodes and malicious nodes present in the network .Such nodes degrade the performance of the network. A survey of reputation based mechanism and credit based mechanism is done which includes different strategies by which non cooperative nodes are detected, isolated and/or prevented, their advantages and limitations. Also, a global reputation based scheme is proposed for the detection and isolation of selfish node. A cluster head is used which is responsible for reputation management of each node in the network. Detection of selfish nodes is accomplished which are created due to nodes conserving their energy. After their detection, performance analysis of network with selfish node and the network after isolation of selfish node is carried out using NS2.

Keywords*-Manet, DSR, Selfish node, reputation based mechanism, ns2.*

I. INTRODUCTION

A Mobile Ad-hoc network (MANET) is a self configuring and infrastructure less network of mobile nodes. Each node acts as a router and is free to move independently in any direction. In an ad-hoc network, communication between two nodes beyond the transmission range relies on intermediate nodes to forward the packet. The communication between nodes takes place using routing protocol which is of three types: Proactive, Reactive and Hybrid routing protocol. Pro-active (table-driven) routing: This type of protocols, such as DSDV maintains fresh lists of destinations and their routes by periodically distributing routing tables throughout the network. The main disadvantage of such algorithms is slow reaction on restructuring and failures.

Reactive (on-demand) routing: This type of protocols, such as DSR, AODV finds a route on demand by flooding the network with Route Request packets. The main disadvantages of such algorithms are high latency time in route finding and network clogging due to excessive flooding.

Hybrid routing: This type of protocols combines the advantages of both proactive and reactive routing. The routing is initially established with some pro actively prospected routes and then serves the demand from additionally activated nodes through reactive flooding. The choice for one or the other method requires predetermination for typical cases. The main disadvantage of such algorithms is that it takes long time when exploring new routes without a prior knowledge.

The reactive routing protocol i.e., DSR protocol can react to topological changes rapidly. Each node gathers information about the network topology by overhearing other nodes' transmissions. This is known as promiscuous mode of operation. DSR is a reactive routing protocol. There are two main operations in DSR; route discovery and route maintenance. DSR protocol tries to minimize the energy consumption by discovering routes to other nodes only when they are required. Each node maintains a route cache to remember routes that it has learnt about. All routing protocols including DSR assume that all nodes in a network are cooperative and forward others' messages.

The successful operation of MANET is totally dependent on the cooperation of participating nodes in communication. Lack of fixed infrastructure in ad hoc networks forces ad hoc hosts to rely on each other in order to maintain network stability and functionality. But sometimes nodes do not work as they were expected to and give rise to reluctant and/or malicious nodes.

In DSR protocol, the selfish nodes are detected using promiscuous overhearing of neighboring node when node drop packet due to nodes conserving their energy. Also, using reputation value and energy value of each node placed at cluster head, selfish node isolation is carried out. Simulation analysis of network is carried out using NS2. Along with the detection and isolation of selfish node using global reputation, this paper also gives a review on various types of reputation based and credit based mechanisms by which selfish and malicious nodes are detected, isolated and prevented.

II. TYPES OF NON COOPERATIVE NODES

In an ad hoc network, the transmission range of mobile nodes is limited due to power constraint. Hence communication between two nodes beyond the transmission range relies on intermediate nodes to forward the packets. But sometimes these intermediate nodes do not work as expected, in order to conserve their limited resources such as energy, bandwidth etc. Such nodes are called non cooperative nodes or misbehaving nodes. They are of following types:

Malicious Nodes: If malicious nodes are present in a MANET, they may attempt to reduce network connectivity by pretending to be cooperative, but in effect drop any data they are meant to pass on. Several types of attacks are performed by malicious node like DOS attack, black hole attack, worm hole attack, rushing attack. The attacks of malicious node on other nodes could be in the form of unnecessary route request control message, frequent generation of beacon packets or forwarding of stale information to nodes. These actions may result in defragmented networks, isolated nodes, and drastically reduced network performance.

Selfish Nodes: Selfish nodes work in an ad hoc network to optimize their own gain, with neglect for the welfare of other nodes. Selfish nodes disturb the performance of ad hoc network to a great extent. When a node becomes selfish it does not cooperate in data transmission process and causes a serious affect on network performance. It simply does not forward packets of other nodes to conserve its own energy, bandwidth. Selfish nodes can be divided into two categories:

Category I: The nodes participate correctly in routing function but not forward data packets they receive to other nodes; so data packets may be dropped instead of being forwarded to their destination.

Category 2: The nodes do not participate correctly in routing function by not advertising available routes. For example: in DSR, selfish nodes may drop all RREQ they receive or not forward a RREP to some destination. Consequently, these selfish nodes will not participate in the requested routes.

III. RELATED WORK

Most of the existing solutions are based on following mechanisms: reputation based, credit based and Reputation cum Credit based mechanism.

3.1. REPUTATION BASED MECHANISM

In Mobile Ad hoc network, Reputation systems are used to keep track of the quality of behavior of other nodes. Basically reputation is an opinion formed on the basis of watching node behavior by direct and/or indirect observation of the nodes, through route or path behavior, number of retransmissions generated by the node, through acknowledgement message and by overhearing node's transmission by the neighboring nodes.

One of the main goals/reasons for using reputation systems in a network of entities interacting with each other is to provide information to help assess whether an entity is trustworthy. This helps in detection of selfish and malicious nodes. Another goal is to encourage entities to behave in a trustworthy manner, i.e. to encourage good behavior and to discourage untrustworthy entities from participating during communication.

A mechanism called Watchdog for the detection of non cooperating nodes, and Pathrater for rating of every used path are proposed. The watchdog mechanism is employed on each node individually to observe the message sent by neighboring nodes. Comparison of the overheard messages with a list of messages that have to be forwarded reveals whether the observed node is forwarding the messages appropriately or not. This enables nodes to avoid non cooperative nodes in their routes. The limitation of this mechanism is that the misbehaving node gets isolated, so this becomes reward for misbehaving node and its sole intention of energy saving is accomplished. This algorithm can only detect the misbehavior but unable to do anything to correct it.

CORE, a collaborative reputation mechanism, also has a *watchdog* component; however it is complemented by a reputation mechanism that differentiates between subjective reputations (observations), indirect reputation (positive reports by others), and functional reputation (task specific behavior), which are weighted for a combined reputation value that is used to make decisions about cooperation or gradual isolation of a node.

CONFIDANT protocol uses reputation mechanism to identify and isolate selfish nodes. The protocol is based on selective altruism and utilitarianism, thus

making misbehavior unattractive. CONFIDANT consists of four important components - the Monitor, the Reputation System, the Path Manager, and the Trust Manager. They perform the vital functions of neighborhood watching, node rating, path rating, and sending and receiving alarm messages, respectively. Each node continuously monitors the behavior of its first-hop neighbours. If a suspicious event is detected, details of the event are passed to the Reputation System. Depending on how significant and how frequent the event is, the Reputation System modifies the rating of the suspected node. Once the rating of a node becomes intolerable, control is passed to the Path Manager, which accordingly controls the route cache. Warning messages are propagated to other nodes in the form of an *Alarm* message sent out by the Trust Manager.

Self policing MANET, combines misbehaviour detection method with reputation system. Here each node can make its own decision on how to react to the behaviour of other nodes. Self policing provides a disincentive for cheating by excluding node from network. In this paper, author enhances CONFIDANT protocol and maintains two rating to make decision about the node: reputation rating and trust rating

COSR (Cooperative on Demand Secure Routing Protocol), is an extension of DSR protocol that uses reputation model to detect malicious and selfish behavior of nodes and makes all nodes more cooperative. In COSR, Node reputation and Route reputation are measured using three parameters: *contribution of node* (how many route as well as data packet are forwarded between nodes), *capability of forwarding* packet of a certain node using energy and bandwidth threshold and *recommendation* which represent other's subjective recommendation. Advantage of COSR is that it is capable of avoiding hot points .It work well with blackhole, wormhole, rushing attack and selfish node but unable to handle DOS attack.

However, there are limitations of reputation based mechanism. First, as there is a possibility of collision, a packet will naturally drop even in the absence of a selfish node. This makes it difficult to ascertain whether the packet drop is due to natural reasons or selfish behavior of node. Second, the selfish nodes isolated from the network using reputation based scheme cannot be used in data forwarding. This solution is trivial, but not efficient. Much approach does not punish nodes that do not cooperate since data is forwarded using a different path without complaint. Another limitation of reputation based system is that it often assumes that nodes that send reputation information about their peers are themselves trustworthy; and they are subject to collusion among nodes that misreport reputation information.

3.2. CREDIT BASED MECHANISM

Credit based mechanisms reward nodes for forwarding by giving those credits. Without credit, a node cannot transmit self-generated data packets.

SPIRITE, an incentive based system in which selfish nodes are encouraged to cooperate. In this system, a node reports to the Credit Clearance Service, the messages that it has received/forwarded by uploading its receipts. Intermediate nodes earn credit when they forward message of others' node. In addition to the availability of central authority, sprite assumes source routing, and a public key infrastructure.

Limitations of this mechanism are, a virtual bank is required to manage credits and when a node has enough credits to send its own data, it can decide not to cooperate anymore and starts dropping packets. Also securing messages containing credits is also an essential requirement so that malicious node could not change credit value.

3.3. REPUTATION CUM CREDIT BASED SYSTEM

Secure and Objective Reputation-based Incentive (SORI) scheme encourages packet forwarding and disciplines selfish behavior in a non cooperative ad hoc network.ARM selects low mobility nodes as reputation management nodes and is responsible for managing reputation values. ARM uses locality aware Distributed Hash Table for efficient reputation information collection and exchange. Advantage of using ARM is that ARM builds a hierarchical structure to efficiently manage the RVs of all nodes, and release the reputation management load from individual high mobility nodes. This enables low overhead and fast global reputation information accesses. Also ARM does not require currency circulated in the system.

From above literature survey, following issues will be considered to make comparison on different mechanism.

3.1.1.DETECTION OF NON-COOPERATIVE NODE

Both reputation based system and credit based system uses one of the following technique for the detection of non cooperative node. Promiscuous mode is used to overhear the communication of its neighboring node. In CORE, nodes do not only rely on promiscuous mode, but in addition they can judge the outcome of a request by rating end to end connection. In monitor mechanism is used and neighbor watch mechanism is used. Retransmission of messages, route reply messages and history or previous observations is

also used by different authors to detect non cooperative nodes.

3.1.2. MANAGEMENT DEVICES

Some Reputation and Credit based mechanisms require extra management device or node for the management of reputation or credit. SPIRITE uses CCS for its credit management; ARM uses low mobility devices for reputation management. Other parameters to choose management nodes are high energy, locality and reputation table. This paper uses Cluster head as reputation manager.

3.1.3.ROBUSTNESS AGAINST NON-COOPERATIVE NODE

Systems like CONFIDANT, COSR are robust against non-cooperative node which system like CORE, SORI, ARM work well with selfish node.

3.1.4. ROBUSTNESS AGAINST COLLISION

SPIRITE, CONFIDANT is collusion resistant system.

3.1.5. AUTHENTICATION MECHANISM

SPRITE uses cryptographic method and digital signature to prevent data from malicious node. The propagation of reputation is computationally-efficiently secured by a one-way-hash-chain- based authentication scheme. Utilize hash chains to reduce number of digital signature operation.

3.1.6. GLOBAL/LOCAL REPUTATION

From above references it is carried out that reputation or credit value is kept either globally or locally. Each has advantage as well as disadvantage. In global Reputations each node maintains reputation values of every other node, so the size is O(N) while in Local Reputation each node maintains reputation values of the neighbor node that is located in one-hop. Global reputation needs an additional computational overhead to decide whether to accept or reject a warning message and to update the reputation table. Local reputations are less vulnerable to false accusations than global reputations because it uses direct observation. Global reputation are less reliable as message traverse across the network so it could be delayed, modified, replayed or accidentally lost during transmission. Global reputation has better performance with respect to the mobility issue, because every node knows the behavior of other node in the network so possibility to cheat is less.

IV. PROPOSED SCHEME

This section represents the basic scheme of reputation based isolation of selfish node. The network architecture in figure 1 of proposed scheme consists of n number of mobile nodes and a cluster head. In comparison to the previous mechanism, this scheme uses cluster head as a reputation manager. The advantage of using cluster head is that if it fails, a new cluster head take the responsibility.

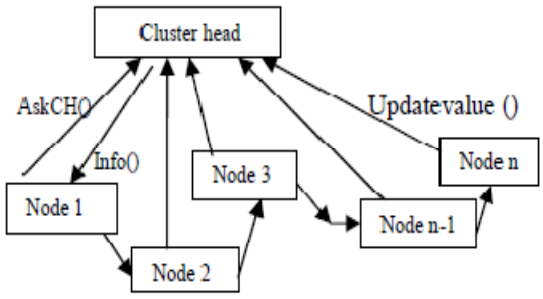


Figure 1: Architecture of proposed scheme

Assumptions: Some assumptions are considered for the proper operation of the scheme:

1. Energy threshold and reputation threshold are assumed as a fixed integer value.
2. It is assumed that cluster head fulfill all the criteria related to degree, location and association with other nodes in the network. It has sufficient energy and it does not misbehave.
3. Each node operates in a promiscuous mode, i.e., each node listens to every packet transmitted by its neighbours even if the packet is not intended for the node.
4. The parameters of each node in the network are almost same. For example, their transmission ranges and energy.
5. Nodes which want to send data already know the path to destination.

Proposed scheme works in three steps. Getting energy and reputation value of each node, Detection of selfish node and Isolation of selfish node from the network.

Let each node have fixed amount of initial energy NE and reputation value R. During the communication of packet, each node consumes a fixed amount of transmission energy (TE) and receiving energy (RE) consequently. At the instance where node energy drops below a pre defined threshold (E_THRESH), the node turns selfish, and drops all packets received from its neighboring nodes. Now if intermediate node forward

packet correctly to its neighboring node, its reputation is increased by one else reputation value is decrease by one. If reputation of any node is less than a pre defined threshold (R_THRESH), node becomes selfish.

The value of each node's energy and reputation is kept at cluster head database table called ER list as shown in table 1.

Table 1: ER list consist of following information

Node ID	Node Energy	Reputation value

Where, Node ID is a unique id of each node.

The value of energy and reputation is updated through a small message called updatevalue containing values (nid, energy, reputation). Each time a node sends other's message to its neighboring node, it forwards *updatevalue()* message to the cluster head for updating energy and reputation values in ER list.

At route discovery phase, each time a node wants to send its packet to other node, it first communicates using *AskCH*() with the cluster head , that knows about the node energy and reputation value of each intermediate nodes present in the path. *AskCH((sn_id, dn_id,int_nid (1,2,….)), r)*

Where *AskCH*() is used to get value from clusterhead, *sn_id* is source node id , *dn_id* is destination node id , *int_nid* contain id of intermediate nodes . *r* is a random number key which is used between clusterhead and source node for encryption and decryption of message, so that no other node is capable of changing the message, thus preventing the message or data from different attacks from malicious node.

Cluster head sends a message

Info((sn_id, dn_id,int_nid (1.1,2.2,….)), r)

Where 1.1, 2.2 and so on contain value of energy and reputation of respective intermediate nodes with r.

If any node is found having low energy value and low reputation value, it is considered as selfish node. If selfish node is present in the path, isolation of such node is carried out by not appending the node in the path. Hence no packet is forwarded through that node and another path is chosen by the sender node.

Global reputation based approaches are considered less reliable since the transmission of packets in or across the network makes them susceptible to be delayed; modification, replay as well as accidental lose during their transmission. For this reason a security mechanism should be applied to the message.

V. SIMULATION SETUP

The performance study of selfish node has been done using NS-2 simulator. NS-2 is a scalable simulation environment for wireless network systems.

The network which is used for simulation consists of 20 nodes placed randomly in 670x670 areas. Each node has a transmission range of 250m and moves at a speed of 10m/s. The total sending rate of all the senders of the multi-cast group, i.e. the traffic load is 1Mbps.

To assign the value of node energy, energy model is used. Every node has initial energy set to 1000 joules. Receiving and transmitting power of node is set as 1 watt.

For performance study of network, two different numbers of connections between nodes were chosen using different values in traffic generator with given simulation parameter as shown in Table 2.

Table 2: Simulation Parameters

Parameter	Value
Number of Nodes	20
Routing Protocol	DSR
Packet size	512 bytes
Traffic model of sources	Constant bit rate
Mobility model	Random way point
Max speed	15 m/s
Initial energy of node	1000 joules
Simulation time	25 sec

From above parameter two different cases are observed. In case I, two selfish nodes are detected and in case II, three selfish nodes are detected.

VI. SIMULATION RESULT

Simulation analysis is carried out using NS2. In this scenario, the reluctant node will drop every incoming packet if that packet is neither from itself nor to itself. To overcome with reluctant node problem, proposed scheme is applied to improve the network performance.

To analysis the network performance, a comparison is made on the basis of node throughput and packet delivery ratio between a network with selfish node and that after isolation of the selfish node using the proposed scheme.

Node Throughput is the average rate of successful message delivery over a communication channel. This data may be delivered over a physical or logical link, or pass through a certain network node. The throughput is usually measured in bits per second (bit/s or bps).

Figure 2 and figure 3 shows throughput of the network with two and three selfish nodes present in the network respectively and the throughput of the network after isolation of network.

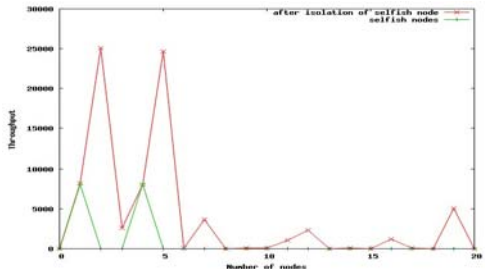


Figure 2: Throughput of ideal network and network with two selfish nodes

From figure 2 it is clear that the throughput gets degraded in the presence of selfish nodes. As selfish nodes increases in the network the performance of the network degrades. When there are 2 selfish nodes present in the network throughput degrades by 80% and when there are 3 selfish nodes in the network throughput get degrade by 90%. Hence selfish nodes should be isolated from the network using the proposed scheme.

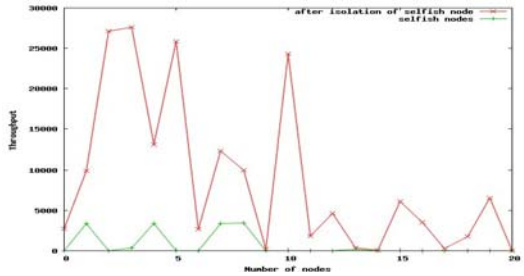


Figure 3: Throughput of ideal network and network with three selfish nodes.

Packet Delivery Ratio (PDR) is the ratio of total no. of packets sent to total no. of packets received.

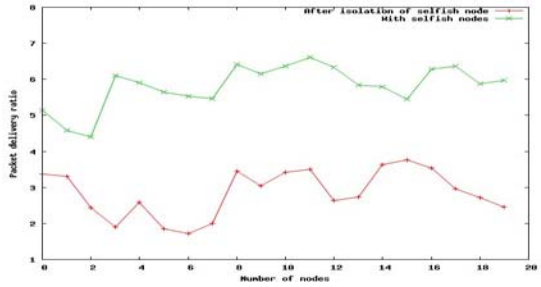


Figure 4: PDR of Network with 2 Selfish nodes and after isolation of Selfish nodes

Figure 4 and figure 5 shows the PDR when there when there are 2, 3 selfish nodes present in the network respectively and the PDF after isolation of selfish nodes.

It is analyzed from the data obtained by trace file that when there are 2 selfish nodes present in the network PDR increases by 50.8% and when there are 3 selfish nodes present in the network PDR increases by 65.2% as compared to the PDF after isolation of selfish node.

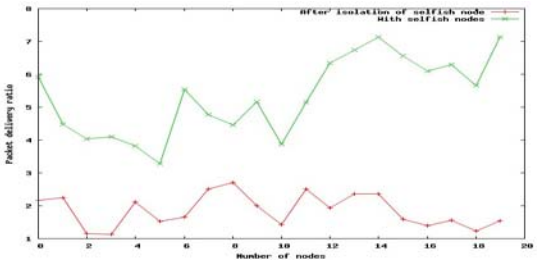


Figure 5: PDR of Network with 3 Selfish nodes and after isolation of Selfish nodes

VII.CONCLUSION AND FUTURE WORK

Here, the study of various reputations based and credit based mechanism which includes issues such as detection and isolation of non cooperative node, authentication mechanism, robustness, and management devices. Using these issues comparison of these mechanisms is done. These mechanisms help nodes to cooperate. With the insight gained from such an understanding, this paper proposes a new scheme based on node energy and reputation value to detect and isolate selfish node. Proposed scheme uses cluster head for keeping the value of energy and reputation of each node in a table. Security mechanism is also applied using cryptographic key so that message can be prevented from malicious node.

Proposed scheme is simulated using NS2.Performance evaluation of the scheme has been carried out which shows that the network throughput and packet delivery fraction increases after applying the proposed scheme.

For future work, the identity of the node can be hashed to further enhance the security. Also, malicious nodes may be taken into consideration**.**

REFERENCES

[1] Fei Wang, Yijun Mo, Benxiong Huang,"COSR: Cooperative on Demand Secure Route Protocol in MANET", IEEE ISCIT, China, pp 890-893.

[2] S. Zhong, J. Chen, and Y. Yang, "Sprite: a simple, cheat-proof, creditbased system for mobile ad-hoc networks," IEEE INFOCOM, San Francisco, CA, USA, Vol 3, pp 1987-1997.

[3] Chee wah Tan,"Enforcing cooperation in an ad hoc Network using cost-credit based forwarding

and Routing Approach", WCNC, IEEE, pp 2935-2939.

[4] Qi He, Dapeng Wu, Pradeep Khosla,"SORI: A Secure and Objective Reputation-based Incentive Scheme for Ad-hoc Networks", WCNC / IEEE Communications Society, Vol. 2, pp 825-830.

[5] Haiying Shen and Ze Li,"ARM: An Account-based Hierarchical Reputation Management System for Wireless Ad Hoc Networks ,The 28th International Conference on Distributed Computing Systems Workshops, IEEE, pp 370-375.

[6] Hameed Janzadeh, Kaveh Fayazbakhsh, bahador bakshi,"A secure credit-based cooperation stimulating mechanism for MANETs using hash chains", Future Generation Computer Systems - Elsvier ,pp 926-934.

[7] Rekha kaushik, Jyoti Singhai "Simulation Analysis of Node Misbehaviour in an Ad hoc Network using NS2 " International journal of computer science and network security, Vol 8 , pp 179-182.

[8] A.V. Babu , Mukesh Kumar Singh "Node Isolation Probability of Wireless Adhoc Networks in Nagakami Fading Channel" International journal of computer networks and communications, Vol 2, pp 21-36.

[9] D.Johnson,Y.Hu,D. Maltz,"The Dynamic Source Routing protocol (DSR) for Mobile Ad hoc network", RFC 4728.

[10] S. Buchegger and J-Y. Le Boudec,"Performance Analysis of the CONFIDANT Protocol:Cooperation Of Nodes, Fairness In Dynamic Ad-hoc Networks", Proc. of the IEEE/ACM Symposium on Mobile Ad Hoc Networking and Computing (MobiHOC).

[11] Sonja Buchegger, Jean Yves Le Boundee, "Self - policing in Mobile Ad hoc Networks" In CRC Press, Chapter Handbook on Mobile Computing,.

[12] NS2 NetworkSimulator. http://www.isi.edu/nsnam/ns

◈◈◈

PMSM Electric Drive System simulation for Parallel Hybrid Electric Vehicle

M. Lakshmi Swarupa[(1)] , G. Tulasi RamDas[(2)] & P.V. Raj Gopal[(3)]

[(1)]Assoc.Prof, Malla Reddy Engg college ,[(2)]Vice-chancellor for JNTUK &[(3)]AGM, BHEL R&D
E-mail : swarupamalladi@gmail.com , das_tulasiram@yahoo.co.in & pv_rajgopal@bhelrnd.co.in

Abstract - As permanent magnet synchronous motor (PMSM) has the characteristics of high power density, high efficiency, and high torque/ current units, so it is extensively used in electric vehicles/Hybrid electric vehicles driving. Energy management unit(EMU) in hybrid electric vehicle (HEV) demands for good motor driving and energy regeneration capability. Based on the mathematical model of PMSM, this paper presents the PMSM vector control algorithm, studies HEV energy regeneration requirements for motors and analyzes power generation work mechanism of PMSM. At last, combined with HEV control strategy, simulation of PMSM system energy regeneration research is done and the simulation results show that the system has good performance.

Keywords- *Permanent magnet synchronous machine, field-oriented control, reference model, PI, NN controllers.*

I. INTRODUCTION

Recently, as the world resources and environmental issues are becoming increasingly prominent, many countries have stepped up the development plans of electric vehicle (EV) [1]. Hybrid electric vehicle (HEV) and the fuel cell vehicle (FCEV) are the new focus. Motor is the core component in HEV driving. In pure electric vehicles and fuel cell electric vehicles, it is the only driving component. In HEV, it is the key component to realize different control strategies. Motors for EV require high reliability, wide torque and speed scope and large power to volume ratio.

The development of permanent magnetic materials, especially the rare earth material, makes the motor power density and torque increase and the volume decrease.

PMSM is widely used in high efficiency, high power density, high torque/current units and wide-speed operation EV driving system. Energy management unit in HEV controls the energy flow according to the work states of the vehicle, this requires the motor has a good driving system and energy regeneration capability. Researching on motor energy regeneration is of great significance for improving energy efficiency and increasing system availability. Based on the PMSM mathematical model, this paper presents the PMSM vector control algorithm, studies HEV energy regeneration requirements for motors, and analyzes power generation work mechanism of PMSM. Then the control strategy of PMSM energy regeneration is proposed and simulation of the motor control system is made .that the PMSM had its permanent magnets mounted on the surface of the rotor. This type of PMSM has therefore a uniform air gap and no saliency, hence Ld = Lq. In the actual demo, it is assumed that the PMSM has an interior permanent magnets rotor.

The impact of the buried-magnet configuration is rotor saliency that makes Lq>Ld and introduces a reluctance torque term into the PMSM torque equation. To take advantage of the reluctance torque, the Id current component is no longer set to zero has it is for the PMSM with surface mounted permanent magnets.

The **PM Synchronous Motor Drive** is composed of four main parts: The electrical motor, the Three-phase Inverter, the VECT controller and the Speed Controller.

- The electrical motor is a 288 Vdc, 100 kW PMSM. This motor has 8 pole and the magnets are buried (salient rotor's type).
- The Three-phase Inverter is a voltage source inverter, controlled by PWM. This block is built using the Universal Bridge Block.
- The VECT controller block computes the three reference motor line currents corresponding to the flux and torque references and then generates a corresponding PWM using a three-phase current regulator. When the nominal flux is required, an optimal control is used in order to minimize the line

current amplitude for the required torque. When a flux weakening is needed, the amplitude and the phase of the current are changed to extend the torque-speed operating range.

- The Speed Controller is used in torque regulation mode. The normalized flux value is computed with the speed of the machine in order to perform a flux weakening control.

The Torque limitation block is used to prevent the limitation due to the torque-speed characteristic of this motor for a 288 Vdc source. When the internal machine's voltage reaches the inverter voltage (because the desired torque is too high for the motor's speed), the inverter becomes in saturation mode (the desired current cannot flow anymore into motor). After this point, there will be a loss of current tracking which will decrease the motor current. This block is used to reduce the reference torque as a function of the motor's speed and the torque-speed characteristic in order to never operate in inverter saturation mode.

II. COMPARISON BETWEEN INDUCTION MOTOR AND PMSM

A summary of the comparison of the inverter fed im and pmsm is given below.

1. **Harmonic current:** Harmonic current for a given harmonic voltage is higher in the case of IM. As result, torque pulsations will be higher.
2. **Switching frequency:** For a given hysteresis band, switching frequency is high in case of IM compare to PMSM. This results in higher inverter losses and low overall drive efficiency in the case of IM.
3. **Power factor:** PMSM operates at almost UPF for all loads, whereas in IM the power factor varies from 0.23-0.72.
4. **Torque:** PMSM operates in true synchronous mode (as long as torque generated is more than the load torque), the torque ratio of PMSM and IM are found to be 2. Higher ratios are possible for higher stator currents. The magnitude of current is limited only by the thermal limitation of the motor in the case of PMSM. This limit is much less in IM due to rotor copper losses.
5. **Windage losses:** Since PM rotor has a smooth surface which is better than the IM, the windage losses are less or equal.
6. **Efficiency:** Higher in case of PMSM compared to IM.
7. **Experimental setup:** Almost the same for IM and PMSM.
8. **Speed encoder:** In the case of IM, since the requirement is only to measure the speed, a simple speed encoder can be used. Whereas in PMSM, a sophisticated encoder can be used, as the encoder pulses are needed to generate the reference signals.
9. **Stress on the switching devices:** Since the IM always operate at lagging PF it draws more current for a given load compared to PMSM. As a result, stress on the switching device is more. PMSM can be operated in leading PF.
10. **Software complexity:** Since slip speed has to be determined in the case of IM, software structure is complex and the time taken by the controller to execute the software program once is higher in IM compared to PMSM.
11. **Response time:** Better in PMSM.
12. **Starting capability:** Good in case of IM. PMSM starts only at low frequency when operated in open loop. This capability can be improved significantly by operating in self-controlled mode.
13. **Stability:** IM is more stable compared to PMSM when operated in open loop.
14. **Reliability:** Aluminum and iron have different coefficients of thermal expansion. This leads to reliability problem over the life of IM if it always operated on full load continuously, or when it is used in applications which require load changes over the cycle. Absence of these windings make PMSM more reliable.
15. **Cost:** Since magnet cost is high, PMSM is very expensive.
16. **Fabrication complexity:** Special jigs are required to fabricate the PM rotor. Assembly of PMSM is more complicated compared to IM.
17. **General comment:** IM are more sensitive to both supply voltage and frequency variations. Torque capability of these motors varies with the square of the voltage, inversely with the square of the frequency. But their speed is less dependent on voltage and directly related to frequency. Whereas in PMSM, the speed is independent of voltage.

III. MATHEMATICAL MODEL OF PMSM

Vector control is one of the control technologies with high performance for PMSM [3]. For the control system, to decide control strategy should be based on specific applications and control purposes. When starting the engine in HEV, the motor should provide constant torque to start engine quickly. In this paper, the maximum torque control strategy is used. To minimize stator current while meet torque demand, this reduces

the copper loss of the motor, and is in favor of inverter switching device to work, but also to reduce costs of the PMSM driving system.

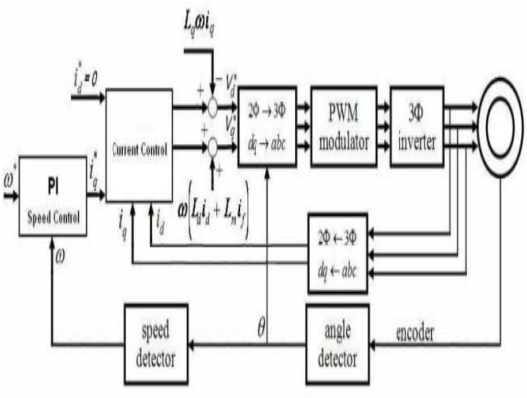


Fig 1: PMSM Control system Block

Using maximum torque/current control, the motor Current vector must meet:

$$\begin{cases} \dfrac{\partial(T_{em}/i_s)}{\partial i_d}=0 \\ \dfrac{\partial(T_{em}/i_s)}{\partial i_q}=0 \end{cases} \quad (1)$$

Also, the following formula can be obtained:

$$T_e = P(\psi_d i_q - \psi_q i_d) = P[\psi_r i_q + (L_d - L_q) i_d i_q] \quad (2)$$

$$i_s = \sqrt{i_d^2 + i_q^2} \quad (3)$$

Substituting equation (3) into (1),

$$i_d = \frac{-\psi_f + \sqrt{\psi_f^2 + 4(L_d - L_q)^2 i_q^2}}{2(L_d - L_q)} = \frac{\psi_f - \sqrt{\psi_f^2 + 4(\rho-1)^2 L_d^2 i_q^2}}{2(\rho-1)L_d} \quad (4)$$

Equation (2) is expressed in per-unit value as follow:

$$T_{em}^* = i_q^*(1 - i_d^*) \quad (5)$$

Where, the basic value of current is:

$$i_b = \psi_f / (L_q - L_d) \quad (6)$$

Basic value torque is:

$$T_b = p\psi_f i_b \quad (7)$$

Substituting the per-unit value of equation (4) to (5), then the relation of direct axis current and torque is:

$$T_{em}^* = \sqrt{i_d^*(1 - i_d^*)^3} \quad (8)$$

$$T_{em}^* = \frac{i_d^*}{2}\left[1 + \sqrt{1 + 4i_q^{*2}}\right] \quad (9)$$

So the stator current i_d^* and i_q^* can be expressed:

IV. DESIGN OF CONTROLLERS

The characteristics of the each of proportional (P), the integral (I), and the derivative (D) controls, and how to use them to obtain a desired response. In this tutorial, we will consider the following unity feedback system:

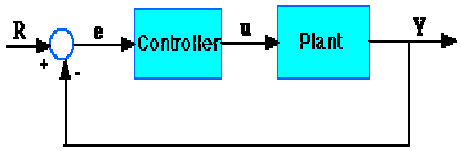


Fig 3: Control system Block diagram

Plant: A system to be controlled Controller: Provides the excitation for the plant; Designed to control the overall system behavior.

The three-term controller:

The transfer function of the PID controller looks like the following:

$$K_p + \frac{K_I}{s} + K_D s = \frac{K_D s^2 + K_p s + K_I}{s} \quad (10)$$

- Kp = Proportional gain
- KI = Integral gain
- Kd = Derivative gain

Working of the PID controller in a closed-loop system using the schematic shown above. The variable (e) represents the tracking error, the difference between the desired input value (R) and the actual output (Y). This error signal (e) will be sent to the PID controller, and the controller computes both the derivative and the integral of this error signal. The signal (u) just past the

controller is now equal to the proportional gain (Kp) times the magnitude of the error plus the integral gain (Ki) times the integral of the error plus the derivative gain (Kd) times the derivative of the error.

$$u = K_p e + K_I \int e dt + K_D \frac{de}{dt} \qquad (11)$$

This signal (u) will be sent to the plant, and the new output (Y) will be obtained. This new output (Y) will be sent back to the sensor again to find the new error signal (e). The controller takes this new error signal and computes its derivative and its integral again. This process repeats.

CLresponse	**Rise time**	**Overshoot**	**Settling time**	**S-s error**
Kp	Decrease	Increase	Small Change	Decrease
Ki	Decrease	Increase	Increase	Eliminate
Kd	Small Change	Decrease	Decrease	Small Change

Table 1: Comparison between controllers

Effects of each of controllers Kp, Kd and Ki on a closed-loop system are summarized in the table shown above. A PI controller responds to an error signal in a closed control loop and attempts to adjust the controlled quantity to achieve the desired system response. The controlled parameter can be any measurable system quantity such as speed, torque, or flux.

The benefit of the PI controller is that it can be adjusted empirically by adjusting one or more gain values and observing the change in system response.

A digital PI controller is executed at a periodic sampling interval. It is assumed that the controller is executed frequently enough so that the system can be properly controlled. The error signal is formed by subtracting the desired setting of the parameter to be controlled from the actual measured value of that parameter. The sign of the error indicates the direction of change required by the control input. The Proportional (P) term of the controller is formed by multiplying the error signal by a P gain, causing the PI controller to produce a control response that is a function of the error magnitude.

As the error signal becomes larger, the P term of the controller becomes larger to provide more correction. The effect of the P term tends to reduce the overall error as time elapses. However, the effect of the P term reduces as the error approaches zero. In most systems, the error of the controlled parameter gets very close to zero but does not converge. The result is a small remaining steady state error. The Integral (I) term of the controller is used to eliminate small steady state errors. The I term calculates a continuous running total of the error signal. Therefore, a small steady state error accumulates into a large error value over time. This accumulated error signal is multiplied by an I gain factor and becomes the I output term of the PI controller.

V. TUNING OF PI CONTROLLERS

Proportional-integral (PI) controllers have been introduced in process control industries. Hence various techniques using PI controllers to achieve certain performance index for system response are presented. The technique to be adapted for determining the proportional integral constants of the controller, called *Tuning*, depends upon the dynamic response of the plant.

In presenting the various tuning techniques we shall assume the basic control configuration wherein the controller input is the error between the desired output (command set point input) and the actual output. This error is manipulated by the controller (PI) to produce a command signal for the plant according to the relationship.

$U(s) = K_p (1+1/ \tau_i s)$(12)

Or in time domain

$U(t) = K_p [e(t) + (1/\tau_i) \int edt]$ (13)

whereK_p = proportional gain

τ_i = integral time constant

If this response is *S*-shaped as in, Ziegler-Nichols tuning method is applicable.

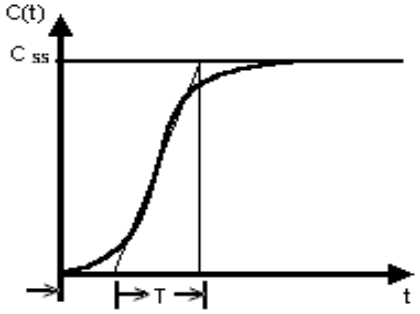


Fig 4: S shaped response of plant

Zeigler- Nichols Rules for tuning PI controllers:

First Rule: The *S*-shaped response is characterized by two constants, the dead time *L* and the time constant *T* as shown. These constants can be determined by drawing a tangent to the *S*-shaped curve at the inflection point and state value of the output. From the response of this nature the plant can be mathematically modeled as

first order system with a time constant *T* and delay time *L* as shown in block diagram.

The gain *K* corresponds to the steady state value of the output C_{ss}. The value of K_p, T_i and T_d of the controllers can then be calculated as below:

KP=1.2(T/L)(14)

τi = 2L(15)

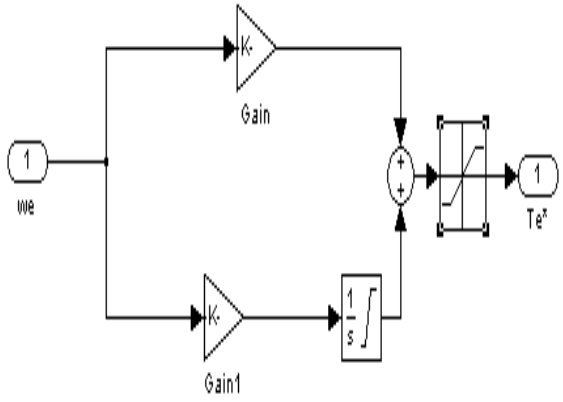


Fig 5: Mathematical Model

VI. NEURAL NETWORKS BASED CONTROLLER

Neural networks can perform massively parallel operations. The exhibit fault tolerance since the information is distributed in the connections throughout the network. By using neural PI controller the peak overshoot is reduced and the system reaches the steady state quickly when compared to a conventional PI controller.

Program for creating the Neural Network:

```
load n
k1=max(i');
k2=max(o1');
P=i'/k1;
T=o1'/k2;
n=157128;
net = newff(minmax(P),[5 1],{'tansig' 'purelin'});
net.trainParam.epochs = 200;
net = train(net,P,T);
Y = sim(net,P);
plot (P,T,P,Y,'o')
gensim(net,-1)
```

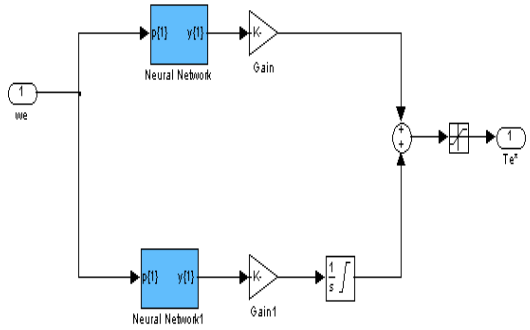


Fig 6: Neural Network based controller

VII. SIMULATION RESULTS

Table 1: Specifications:

J=0.0176	**B=0.0003818**
P=6	**Ld-Lq=0.156**

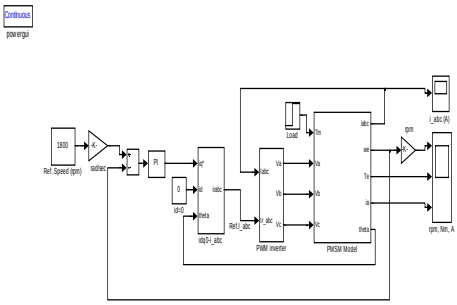
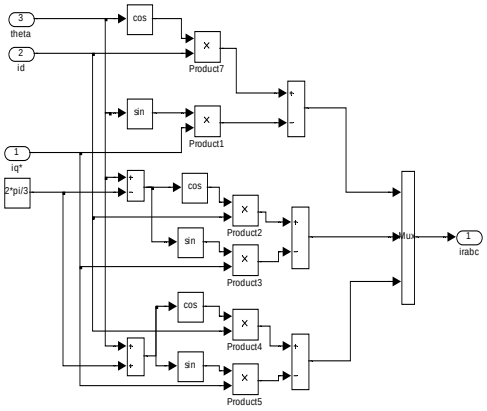
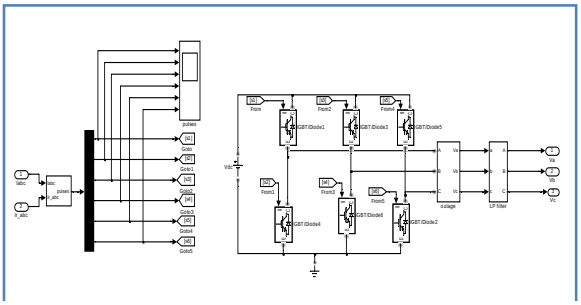
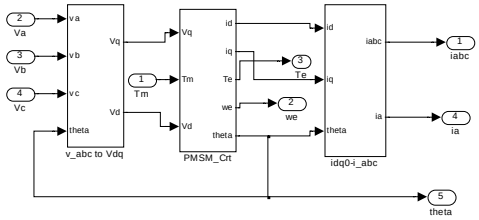
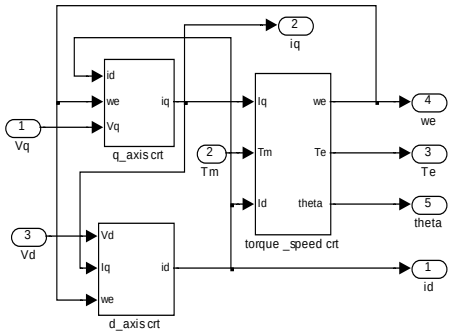


Fig 7: Simulation model of PMSM









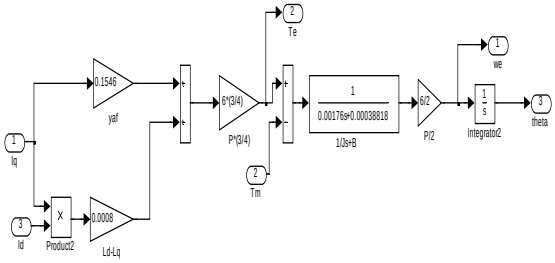
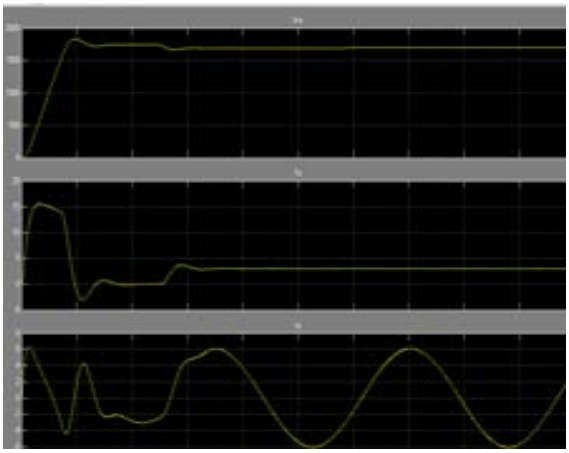
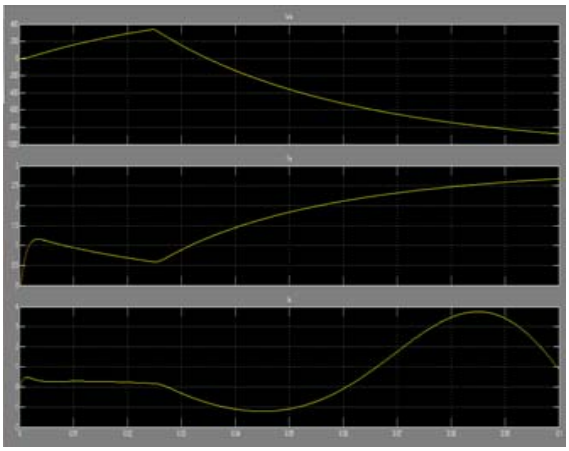


Fig 8: Sub blocks of PMSM



a. PMSM simulation with NN Controller



b. PMSM simulation with PI Controller

Fig 9 (a & b): Simulation Results for NN and PI controller

VIII.CONCLUSIONS

For PMSM control and driving in HEV, according to PMSM mathematical model, this paper deduces PMSM vector control algorithm, studies HEV energy regeneration requirements for motors, and detailed analyzes power generation work mechanism of PMSM. Then the control strategy of PMSM energy regeneration is proposed and the simulation results of the motor control system show that the system has good performance.

ACKNOWLEDGMENT

I would like to first acknowledge and express my sincere thanks to my supervisor& co-supervisor Dr. P.V.RajGopal & Dr. G.TulasiRamDas, for the opportunity that he gave me to work on this highly promising and exciting research area. I would also like to thank Dr.C.K.Sarma, a senior professor from the G.R.I.E.T, whose helped me in understanding mathematical formulation. Finally, a special thank you goes to my parents M.Rama Krishna Rao & M. Girija Kumari for their moral and financial supports and also my husband P.Srikanth for his encouragement throughout.

REFERENCES

[1] Cikanek, S.R.; Bailey, K.E. Regenerative braking system for ahybrid electric vehicle. Proceedings of the American ControlConference 2002 , Volume: 4: 3129 –3134

[2] Wu Hong-xing, Cheng Shu-kang, Cui Shu-mei. Communicationon Vehicle Management Unit in the Electric Vehicle. 12th EML , May 28-31 , 2004 , Saint-Louis , USA. IEEE Transactions onMagnetics, v 41, n 1 II, January, 2005, p 514-517.

[3] Chan, C.C.; Zhang, R.; Chau, K.T.; Jiang, J.Z. Optimal efficiencycontrol of PM hybrid motor drives for electrical vehicles. PowerElectronics Specialists Conference, 1997. PESC '97 Record., 28th Annual IEEE , Volume: 1 , 1997: 363 -368 vol.1

[4] W. Xia , Philip. Chin. A Specially Designed EV PM Motor Drive.Energy Management and Power Delivery, 1998.Proceedings ofEMPD '98. 1998 International Conference on , Volume

◈◈◈

Union of RSA algorithm, Digital signature And KERBEROS in cloud security

Mehdi Hojabri & Mona Heidari

Department of CS and SE Andhra University , Vizag, India
E-mail : hozhabri64@gmail.com, monaheid@gmail.com

Abstract **-** as we know cloud computing is one of the famous part of IT enterprise and nowdays it make many challange for growing the security issues. The security is most important for all domain in computer network and also the cloud is one facility with internet platform. All user, company and enterprise for using the cloud,at the frist they expect more secure,reliable and relation issues. In this paperwe enhance the security issues in cloud computing with mixed three security service such as:

- Kerberos authentication service
- Rsa algorithm
- Digital signature
- Admin

We used the kerberos authentication service for issues the ticket and granting ticket for all users just for make more security. Digital signature with RSA algorithm, to encrypting the data while we are transferring it over the network. A digital signature or digital signature scheme is a mathematical scheme for demonstrating the authenticity of a digital message or document. At the first all users for connecting to the cloud they should make the profile and geting the password from kerberos and after that they can entry to cloud realm for using the cloud and share the information with effect of Rsa algorithm and digital signature. In this article we define a admin.Admin should make the restrict for all user according their IP.

Keywords*: Rsa algorithm , Digital signature , kerberos authentication service,Admin.*

I. INTRODUCTION

A cloud computing facility, paired with technologies such as netbooks, rich internet applications , smart phones, and web services, allows users to run their applications anywhere and access to the desire services at any time. The security in the cloud is one of the most important issues. Already many researcher, survey cloud security problem with rsa algorithm and digital signature. But we add another service for enhance the security issues. This service is kerberos authentication service. In this theory admin define some IP address for using the cloud server provider. It means admin make restrict for some user. At the next step the user with that IP address can connect to the kerberos and after this service it should connect to the cloud service provider for sending the data with influence of digital signature and rsa algorithm. So i think with this long filtering we can enhance the security problem in the cloud.

II. THE CLOUD COMPUTING STACK

We can define cloud Computing as a stackin 3 domain. Below we define tree type of cloudstack:

A. Software as a Service

B. Platform as a Service

C. Infrastructure as a Service

- Software as a Service

software that is deployed over the internet… With SaaS,a provider licenses an application to customers either as a service on demand, through a subscription, in a "pay-as-you-go" model, or (increasingly) at no charge when there is opportunity to generate revenue from streams other than the user, such as from advertisement or user list sales. SaaS is a rapidly growing market as indicated in recent reports that predict ongoing double digit growth. This rapid growth indicates that SaaS will soon become commonplace within every organization and hence it is important that buyers and users of

technology understand what SaaS is and where it is suitable.

- Characteristics of SaaS
- Web access to commercial software.
- Software is managed from a central location .
- Software delivered in a one to many model.
- Users not required to handle software upgrades and patches.
- Application Programming Interfaces allow for integration between different pieces of software.
- Platform as a Service

Platform as a Service (PaaS) brings the benefits that SaaS bought for applications,but over to the software development world. PaaS can be defined as a computing platform that allows the creation of web applications quickly and easily and without the complexity of buying and maintaining the software and infrastructure underneath it.PaaS is analogous to SaaS except that, rather than being software delivered over the web, it is a platform for the creation of software, delivered over the web.

- Characteristics of PaaS:
- Services to develop,test, deploy, host and maintain applications in the same integrated development environment. All the varying services needed to fulfil the application development process.
- Web based user interface creation tools help to create, modify, test and deploy different UI scenarios.
- Multi-tenant architecture where multiple concurrent users utilize the same development application.
- Built in scalability of deployed software including load balancing and failover.
- Integration with web services and databases via common standards.
- Support for development team collaboration – some PaaS solutions include project planning and communication tools.
- Tools to handle billing and subscription management.
- Infrastructure as a Service

Infrastructure as a Service (IaaS) is a way of delivering Cloud Computing infrastructure servers, storage, network and operating systems – as an on-demand service. Rather than purchasing servers, software, datacenter space ornetwork equipment, clients instead buy those resources as a fully outsourced service on demand.

- Characteristics of IaaS
- Resources are distributed as a service
- Allows for dynamic scaling
- Has a variable cost, utility pricing model
- Generally includes multiple users on a single piece of hardware

III. PROBLEM STATEMENT

- Kerberos authentication service

Kerberos is one of secure method for authenticating a request for a service in a computer network. Kerberos was developed in the Athena Project at the Massachusetts Institute of Technology.The name is taken from Greek mythology. Kerberos was a three headed dog who guarded the gates of Hades.Users after send the request,they can take ticket and granting ticket and finally that can be used to request a particular servicefromaserver.Theuser'spassword does not have to pass through the network.

- RSA algorithm

RSA was created by Ron Rivest, Adi Shamir, and Leonard Adleman in 1978. The RSA scheme is a block cipher in which the plaintext and cipher text are integers between 0 and n-1 for some n. A typical size for n is 1024 bits, or 309 decimal digits.That is, n is less than 21024.

Till now, it is the only asymmetric (i.e. needs two different keys) algorithm used for private/public key generation and encryption.RSA is widely used in electronic commerce protocols, and is believed to be sufficiently secure given sufficiently long keys and the use of up-to-date implementations.fig.1

The RSA algorithm involves three steps:

- Key generation
- Encryption
- Decryption
- Digital signature

A digital signature or in simple manner is handwritten signature. Is a authenticates electronic documents. This signature cannot be fake and it asserts that a named person wrote or otherwise agreed to the document to which the signature is attached. The

recipient of a digitally signed message can verify that the message originated from the person whose signature is attached to the document and that the message has not been altered either intentionally or accidentally since it was signed. Also, the signer of a document cannot later disown it by claiming that the signature was forged. In other words, digital signatures enable the "authentication" and "non-repudiation" of digital messages, assuring the recipient of a digital message of both the identity of the sender and the integrity of the message.

IV. DIGITAL SIGNATURE WITH RSA ENCRYPTION ALGORITHM AND KERBEROS TO ENHANCE DATA SECURITY IN CLOUD

In Cloud computing, we have many problem like security of data, files system, backups, network traffic, host security,alter the data and many more.Here at the first we are proposing a concept of kerbros authentication service for issues the ticket for all participating over the network. In this article we suppose any user or client for using the cloud and connecting to cloud, at the fist they must connect to the kerberos and make the profile of information in kerberos data base for more secure.At the next step the AS of kerberos do verifies usere and created the ticket granting ticket and session key and it sent to the users. the next step users send the ticket granting ticket and session key to TGS for get the service. In the next step the TGS send ticket and session key for user.In final step the users send the request service to cloud service provider for using the cloud service and also cloud service, provide service tousers. The process of this scenario is in five step:

A. The client logs on the workstation and send the requests access a ticket-granting ticket on behalf of the user by sending its user's ID to the AS, together with TGS ID, indicating a request to use the TGS service.

B. The AS responds with a ticket that is encrypted with a key that is derived from user password. When this response arrives at the client, the client prompts the user for his or her password, generates the key, and attempt decrypt the incoming message. If the correct password is supplied, the ticket is successfully recovered. Because only the correct user should know the password, only the correct user can recover the ticket. Thus, we have used the password to obtain credentials from kerberos without having to transmit the password in plaintext. The ticket itself consist of the ID and network address of the user, and the ID of the TGS. This corresponds to the first scenario. This is that this ticket can be used by the client to request multiple cloud service granting tickets. So the ticket granting ticket is to be reusable. However,we do not wish an opponent capture the ticket and waits until the user has logged off his or her workstation. The opponent either gain access to that work station or configure his workstation with the same network address as that of the victim.The ticket include a timestamp ,indicating the data and time for which the ticket was issued,and a lifetime, indicating the length of time for which the ticket is valid. Thus,the client know has a reusable ticket and need not bother the user for a password for each new service request.

C. The client request a service-granting ticket on behalf of the user. For this purpose,the client transmits a message to the TGS containing the user's ID,the ID of the desire cloud service,and the ticket-granting ticket.

D. The TGS decrypt the incoming ticket and verifies the success of the decryption by the presence of its ID. It check to make sure that the lifetime has not expired. Then it compares the user ID and network address with the incoming information to authenticate the user. If the uses is permitted access to V, the TGS issues a ticket to grant access to the requested cloud service provider. The service-granting provider ticket has the same structure as the ticket-granting ticket.Indeed,because the TGS is a server,we would expect that the same elements are needed to authenticate a client to the TGS and to authenticate a client to an application server. Again,the ticket contain a timestamp and lifetime. If the user wants access to the same cloud service at a later time,the client can simply use the previously acquired service-granting ticket and need not bother the user for a password. Note that the ticket is encrypted with a secret key(K_v) known only to the TGS and the server, preventing alteration. Finally, with a particular cloud service –granting ticket, the client can gain access to the corresponding service with next step.

E. The user request access to cloud service on behalf of the user. For this purpose the client transmits a message to the server containing the user's ID and the cloud service granting ticket, The server authentication by using the contents of the ticket.

After this process the users can used the cloud service provider, but for making more secure we define the Rsa algorithm and digital signature in the cloud realm. All users at the first must pass the filter of kerberos and the second they can entry to cloud realm, after entry to cloud realm and taking the cloud service, for more secure they must doing follow the Rsa algorithm and digital signature.

Assume we have two enterprises A and B. These enterprise passed the kerberos filtering after that An enterpriseA have a public cloud server with data, applications and many more thing. Company B wants a secure data from A Cloud. We are here, trying to send a secure data to B by using Digital signature with RSA algorithm. We are taking some steps to implementing Digital signature with RSA encryption algorithm.

Suppose we have two employee in this scenario: Ali and Reza.

Step1.Ali takes a document from cloud, which Reza wants.

Step2.The document will crunched into few lines by using some Hash function.Fig.2

Step 3. Ali software then encrypts the message digest with his private key. The result is the digital signature.Fig.3

Step 4. Using RSA Algorithm, Ali will encrypt digitally signed signature with Reza public key and Reza will decrypt the cipher text to plain text with his private key and Ali public key for verification ofsignature.

All this step act according the ticket granting ticket.

Key Generation

Select p, q	p, q both prime, $p \neq q$
Calculate $n = p \times q$	
Calculate $\phi(n) = (p-1) \times (q-1)$	
Select integer e	$\gcd(\phi(n), e) = 1;\ 1 < e < \phi(n)$
Calculate d	
Public key	$KU = \{e, n\}$
Private key	$KR = \{d, n\}$

Encryption

Plaintext:	$M < n$
Ciphertext:	$C = M^e \pmod n$

Decryption

Ciphertext:	C
Plaintext:	$M = C^d \pmod n$

Fig.1 Rsa algorithm

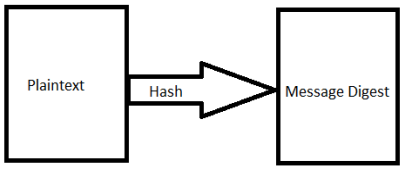


Fig.2 Message Digest

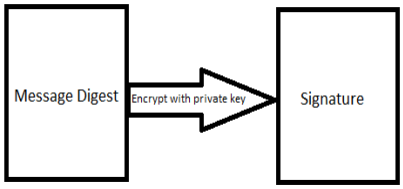


Fig.3 Encryption of message digest into Signature

CONCLUSION

As we read in this article, we survey the problem in cloud service provider. Already some researcher did the enhance security problem in cloud with effect of rsa algorithm and digital signature. But in this paper we investigated the security issues in cloud service with effect of authentication service. I mean we do the combination of three service such as:

- Rsa algorithm
- Digital signature
- Kerberos

We propose the new innovation for improving the security problem in cloud computing. In my opinion in this paper we define one way for enhance the security problem with three filtering. At the first each IP must be confirm with admin.Secont they should apply for taking the ticket.Finally they can catch the cloud service provider,but in the cloud realm they must use the rsa algorithm and digital signature for sending the information.

REFERENCE

[1] Uma Somani, Kanika Lakhani and Manish Mundra "Implementing Digital Signature with RSA Encryption Algorithm to Enhance the Data Security of Cloud in Cloud Computing", 2010 1st International Conference on Parallel, Distributed and Grid Computing (PDGC - 2010).

[2] Kresimir Popovic and Zeljko Hocenski "Cloud computing security issues and challenges", MIPRO 2010, May 24-28, 2010, Opatija, Croatia.

[3] Richard Chow, Philippe Golle, Markus Jakobsson,Ryusuke Masuoka, Jesus Molina Elaine Shi, Jessica Staddon, Ryusuke Masuoka, and Jesus Molina "Controlling Data in the Cloud: Outsourcing Computation without Outsourcing Control",CCSW 09, November 13, 2009, Chicago, Illinois, USA.

[4] (U.S.) Nicholas. Carr, fresh Yan Yu, "IT is no longer important: the Internet great change of the high ground - cloud computing," The Big Switch: Rewining the World, from Edison to Google, CITIC Publishing House, October 2008 1-1

[5] S. Pearson, "Taking Account of Privacy when Designing Cloud Computing Services", CLOUD 09,May 23, 2009, Vancouver, Canada.

[6] M. Casassa-Mont, S. Pearson and P. Bramhall, "Towards Accountable Management of Identity and Privacy:Sticky Policies and Enforceable Tracing Services",Proc. DEXA 2003, IEEE Computer Society, 2003, pp.377-382.

[7] http://csrc.nist.gov/groups/SNS/cloud-computing/.

[8] http://csrc.nist.gov/groups/SNS/cloud-computing/cloud-def-v15.doc

[9] Virtualization – The ability to increase computing efficiency http://broadcast.rackspace.com/hosting_knowledge/whitepapers/Revolution_Not_Evolution-Whitepaper.

[10] Cloud Security Alliance, Security Guidance for Critical Areas of Focus in Cloud Computing V2.1,http://www.cloudsecurityalliance.org/, December 2009

[11] https://incometaxindiaefiling.gov.in/portal/faq_signature.do

[12] Kalyani D. Kadam,Sonia K. Gajre,R. L. Paikrao,Security issues in Cloud Computing

◈◈◈

Channel assignment and rounting in Multi-Channel Wireless Mesh Networks

Geeta Y.Midha & R.K. Krishna

Deptof CSE, Dept of Electronics., R.C.E.R.T., Chandrapur(M.S.), India.
E-mail :geeta_midha@rediffmail.com , rkrishna40@rediffmail.com

Abstract - we examine wireless mesh networks and present a theoretical model for evaluating the expected throughput for fast moving users. We propose intelligent techniques for improving the throughput: distribution of neighbour nodes over the available wireless interfaces and distributed channel assignment for minimising the interference.
Evaluations show that the neighbour nodes should be equally distributed over the wireless interfaces based on the relative difference in hop count to the closest gateway. We also present load-aware techniques which dynamically adapt the nodes according to the current traffic conditions. Load-aware techniques are able to achieve larger throughput but in many cases gain margins are smaller than expected.
It is also shown that a poor choice in neighbour interface binding or channel assignment technique leads to a decreased performance in such away that adding interfaces or channels would not further improve the throughput.
The idea of exploiting multiple channels is particularly appealing in wireless mesh networks because of their high capacity requirements to support backbone traffic.

To reap the full performance potential of this architecture, we develop a set of centralized channel assignment, bandwidth allocation, and routing algorithms for multi-channel wireless mesh networks. A detailed performance evaluation shows that with intelligent channel and bandwidth assignment, equipping every wireless mesh network node with just 2 NICs
operating on different channels can increase the total network goodput .

Keywords: *topologies, interference, channel assignment, multichannel*.

I. INTRODUCTION

A Wireless Mesh Network (WMN) is formed by a set of gateways, mesh routers, and mesh clients. Gateways and mesh routers form the backbone of the network, where mobility is reduced.

Mesh clients can be cell phones, laptops or other wireless devices. Routers communicate with the external network (e.g. the Internet) by forwarding each other's trac (including clients trac) towards the gateway nodes, which are directly connected to the wired infrastructure.

In a WMN, each router forwards packets on behalf of other nodes (that may not be within direct wireless transmission range of their destinations).

Moreover, the gateway functionalities enable the integration of WMNs with various existing wireless networks such as Wi-Fi, cellular networks, WiMax, among others.

Paper is based on the novel definition of co-channel interference which can capture the impact of interference by fully considering both interferenceand connectivity, we define the *Interference-Aware Robust Topology* I-ART) problem which seeks network topology design and a channel assignment such that the induced network topology has the minimum network interference among all 2-connected topologies.

In this work,

2-connectivity is required for survivability and load-balancing purposes. We assume the transmission power of each NIC is fixed. So the topology control problem studied here is quite different from all previous topology control problems in which the network topology is controlled by carefully adjusting the transmission power at each node.

The authors developed a set of centralized algorithms for channel assignment, bandwidth allocation, and routing.

Load-Aware Channel Assignment A central design question in *Hyacinth* architecture is how to bind each network interface to a radio channel.

- Unlike in cellular networks, neighboring nodes in *Hyacinth* communicate over wireless links, and therefore need to share common channels with each other.
- On the other hand, as more NICs in an interference neighborhood share a particular channel, the capacity of each NIC and its associated virtual links reduces.

Therefore, in our algorithm neighboring nodes distribute their NICs across as many channels as possible while maintaining the required connectivity among themselves.

- As some links in the network may carry more load than others, link load information is also considered while distributing the virtual links across different channels.
- At the end, our load-aware channel assignment results in a proportional bandwidth allocation, where more heavily loaded links get more capacity.

Experimental work:

WORK I:

The body of research is to find an optimal channel for a single packet transmission, essentially avoiding inter-ference and enabling multiple parallel transmissions in a neighborhood. [16]

The architecture does not perform channel switching on a packet-by-packet basis; our channel assignment lasts for a longer duration, such as hours or days, and hence does not require re-synchronization of communicating network cards on a different channel for every packet.

This property makes it feasible to implement our architecture using commodity 802.11 hardware. Additionally, our system takes a more global approach by adjusting channel assignments and routes based on the overall network traffic patterns.

The goal of channel assignment in a multi-channel wireless mesh network is to bind each network interface to a radio channel in such a way that the available bandwidth on each virtual link is proportional to its expected load.

A simple approach to the channel assignment problem is to assign the same set of channels to the interfaces of each node, e.g., channel 1 to the first NIC, channel 2 to the second NIC

This *identical channel assignment* indeed provides throughput gains by utilizing multiple channels. Data flow is high, Bandwidth will be high,Packet delay is less ,collision will be less ,and finally sequence no. generated will not be produced .

Single- channel wireless mesh network packet losses in wireless network

There are various causes for packet losses in wireless multihop networks .

We classify the various reasons into three groups

1. Channel induced factors

This includes the random bit error from signal attenuation, multipath fading, shadowing and noise.

2. Interference induced factors

Includes in and out of the mesh network that operates on the same channel as the desired transmission.

3. Node induced factors

For example on the linux system each processor has soft net data structure which holds a list with the incoming packets received by the network interface card(NIC).

ALGORITHM :

Algorithm 1 Interference Aware Robust Topology control[11]

Find 2-connected subgraph of G, G (V, E) such that G has the

minimum number of edges;

for each link e in G **do**

Find the *PIE*(*e*) and calculate the *PIN(e)* of e;

end for

Initialize A(u) to ∅ for all u ∈ V ;

for (all the links in G) **do**

Select links one by one *in a descending order of PINs*;

For the selected link e ∈ G , assign channels for all edges in

PIE(*e*) based on the following rules:

for all end nodes of edges in *PIE*(*e*) **do**

Select the nodes in *PIE*(*e*) one by one *in a descending order*

of node degree;

if there are l(≥ 1) empty NICs on the selected node u

then

Use the l least used channels to fill all the empty NICs on node u;

for all *unassigned edges* e= (u, v) where v has empty NICs **do**

Assign the currently least used one among the l channels to edge e ;

Assign corresponding channel on node v;

end for

for all *unassigned edges* (u, v)where v has NO empty NIC **do**

Channel Swap(G, u, v, (u, v));

end for

end if

if (No empty NIC on node u) **then**

for all *unassigned edges* e= (u, v) where v has empty NICs **do**

Assign the currently least used channels on u on e ;

Assign corresponding channel on node v;

end for

for all *unassigned edges* (u, v) where v has NO empty NIC **do**

Channel Swap(G, u, v, (u, v));

end for

end if

end for

end for

Algorithm 2 Channel Swap(G, u, v, (u, v))

if A(u) ∩ A(v) = ∅ then

doing nothing

else

let k be the least used channel in *PIE*(*e*) among channels in

A(u) ∪ A(v). Without loss of generality, assume that A(u).

Let k = k be an channel in A(v) that is most used in *PIE*(*e*).

Replace k in A(v) by k.

for all the edges (v, w) already assigned **do**

if the change of A(v) makes A(v) ∩ A(w) = ∅ **then**

replace k in A(w) by k.

This replacement may be performed recursively.

end if

end for

end if

Outputs:

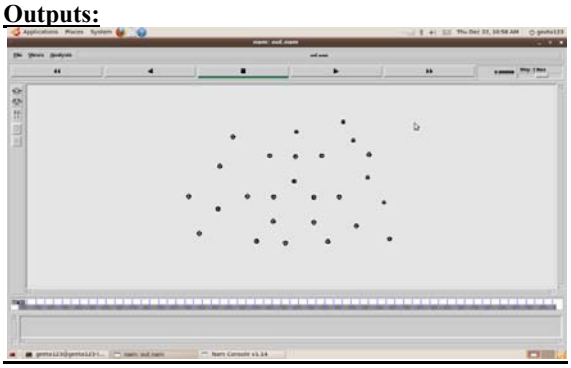


Fig 1:representing nodes

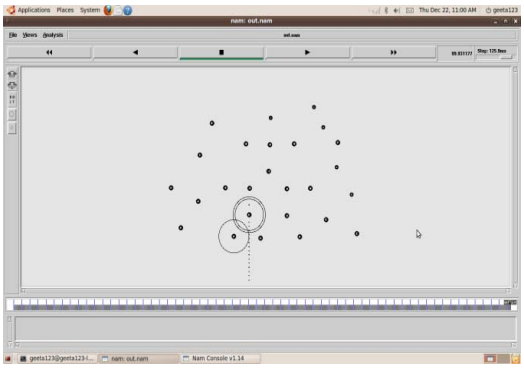


Fig 2: representing nodes with packet dropings

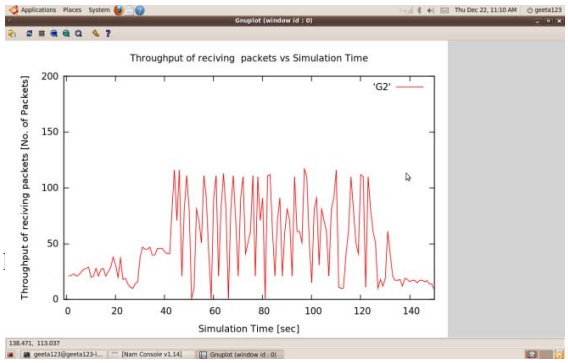


Fig 3: output graph between throughput of reciving packets vs simulation time

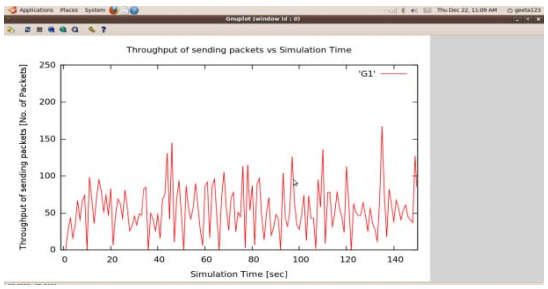


Fig 4: output graph between throughput of sending packets vs simulation time

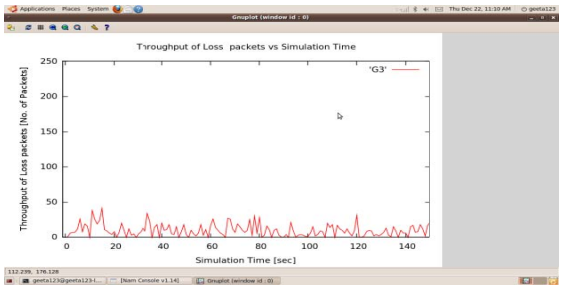


Fig 5: output graph between throughput of loss packets vs Simulation Time

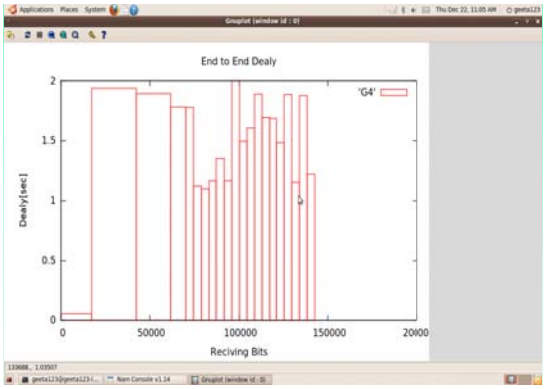


Fig 6: output graph between Delay[sec] vs Reciving Bits

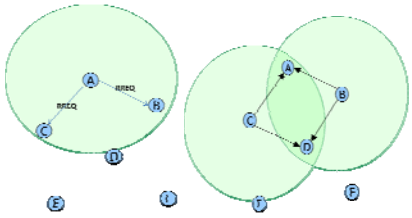
Work II:

AODV-MR has been proposed to Improve AODV to work in WMNs.

AODV-MR assumes that each node has at least one common channel with neighbors. AODV-MR broadcast the Router Request Message (RREQ) on all interfaces.

The first RREQ message received by destination or intermediate nodes is sel-ected and all duplicate RREQs are dis-carded. In addition, AODV-MR increases the network capacity because it causes lower degree of interference and contention time.

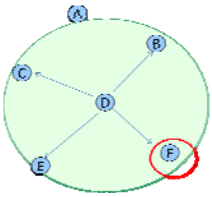
- The CA-AODV has been proposed to assign channel within K hops in ad hoc network.
- Therefore, this allows for concurrent transmission on the neighbouring links along the path and effectively reduces the intra-flow interference.
- AODV routing protocol is a reactive distance vector routing that has been optimized for mobile ad hoc .
- When a source node has a data to transmit, it checks the route table for destination entry.
- In case the route is unknown, it generates and broadcasts route request packets to its neighbours.
- Each RREQ packet has a unique ID which used as identifier and the sequence number indicates the freshness of control packet.
- When an intermediate node receives the RREQ packet that it has not seen before, or it does not has fresh route to destination, a reverse path to destination is created and rebroadcast after incrementing the hop count



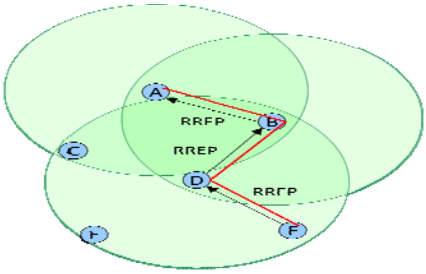
Example: fig 7:RREQ

- Node A needs to communicate with F
- RREQ A->F is released to network
- Neighbors C and B receive RREQ and learn route to A

Intermediate nodes C and B do not have route to F RREQ is broadcasted forward with increased *hop count* only if hop limit is not yet reached A receives it's own RREQ *Reverse paths* to B and C are formed RREQ is discarded Intermediate node D receives multiple copies of RREQ form A Direct routes to C and B are formed. The first arrived RREQ is set used to form route to A (e.g. B here)



- D forwards RREQ,B and C discard duplicate RREQ and learn route to D ,Destination node F finally gets RREQ .



- Route reply packet (RREP) is sent back to node A along reverse route

  In fact any node, which has a fresh route to destination can send RREP and therefore end route search

- Active *forward path* from A to F is created Intermediate nodes also have now active *forward path* to F .

- Route is ready for data transmission.

OUTPUT:

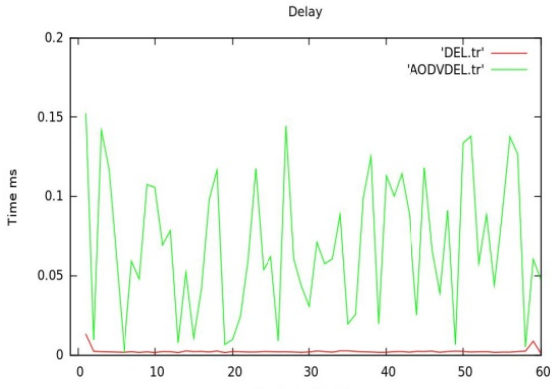


Fig 8: output graph between Time vs Delay

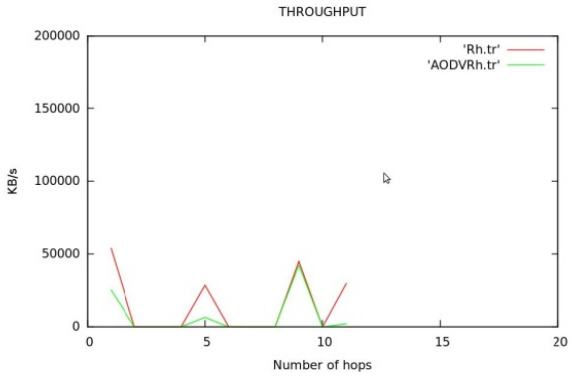


Fig 9: output graph Number of hops in KB/s

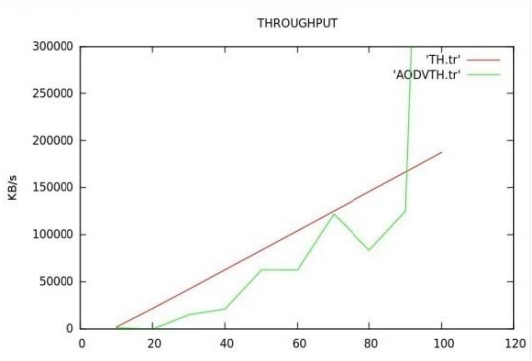


Fig 10: output graph Number of hops in KB/s

Simulation Environment

NS2 is the simulator used for our work. Network simulation software enable us to predict behavior of a large-scale and complex network system such as Internet at low cost Many network simulators, such as NS2, Openet, Omnet, Qualnet, etc., are widely available.

*The stimulation is node movement .Node movement indicates the multichannel interface with other channel.*Node 19 which is indicated by red coloured it means Node 19 communicates the multichannel interface.*Other nodes are for data transfer.

WORK III :

These networks provide relatively high-performance wireless accesses, and are widely adopted at homes, in offices, and hotspot areas in cities. A wireless mesh network is a multihop wireless network consisting of a large number of wireless nodes, some of which are called gateway nodes and connected with a wired network.

It has attracted much research attention due to its potential applications, including last-mile broadband Internet access, neighborhood gaming, Video-on-Demand (VoD), distributed file backup, video surveillance and so on.Due to the limited channel capacity, the influence of interference, the large number of users and the emergence of realtime multimedia applications, improving network capacity have become a critical requirement in such networks.

One common technique used to improve overall network capacity is use of multiple channels. Using multiple channels instead of a single channel in multihop wireless networks has been shown to be able to improve the network throughput dramatically .

Due to path loss effect over the distances ,these long links are lossy and of low throughtput.The

performance of routing protocols can be improved by better defining route metrices and explicitly taking into account of the quality of wireless links .

Second, each wireless nodes is usally euiped with one or more radios that can be switched among multiple,non overlapping channels. Use of multi-radios and multichannels has thus been explored to construct interference –free/mitigated routes on which different channels are Associated with different radios in order to eliminate intra and inter-flow interference.The latter approach has been reffered to as joint routing and channel assignment .

Channel Assignment and routing in Multi-radio, multi-channel Environments

A traditional channel assignment problem is what channel should be assigned to a transmission pair in order to enable transmission, mitigate inter-/intra-interference, and improve network capacity.

This problem is augmented with another dimension in multi-radio and multi-channel environments: what channel should be associated with each of the radio interfaces a node possesses?

Although there have been some preliminary work [10, 11], a rigorous treatment of this problem has been lacking. This problem is further complicated, when it is considered in conjunction with routing.

Several research efforts have been made to address the joint problem of channel assignment and routing, and various heuristics (although with insightful theoretical base) have been proposed under certain (perhaps unrealistic) interference models.

The challenge however, remains to consider the problem in an analytic framework under a realistic interference model (in which cumulative interference due to concurrent transmissions is faithfully characterized). multiradio mesh nodes have the potential to significantly improve the performance of mesh networks, efficient channel assignment is a key issue in guaranteeing network connectivity while still mitigating the adverse effects of interference from the limited number of channels available to the network. A WMN node needs to share a common channel with each of its communication-range neighbors with which it wishes to set up a virtual link 1 or connectivity. However, to reduce interference, a node should minimize the number of neighbors with which it will share a common channel. There is thus a trade-off between maximizing connectivity and minimizing interference.

Common Channel Assignment is the simplest scheme. In this case the radio interfaces of each node are all assigned the same set of channels. The main benefit is that the connectivity of the network is the same as that of a single-channel approach,while the use of multiple channels increases network throughput.

Distributed Load-Aware Channel Assignment :

The neighbor discovery and routing protocolin the previous subsection allows each WMN node to connect with its neighbors and identify a path to the wired network. We now discuss the mechanisms through which a WMN node can decide how to bind its interfaces to neighbors and how to assign radio channels to these interfaces without global coordination, as in the case of centralized algorithm.

Neighbor-Interface Binding:

The key problem in the design of a distributed channel assignment algorithm is channel dependency among the nodes. This channel dependency relationship among network nodes makes it difficult for an individual node to predict the effect of a local channel re-assignment decision.Specically, the set of NICs that a node uses to communicate with its parent node, termed UP-NICs, is disjoint from the set of NICs the node uses to communicate with its children nodes, called DOWN-NICs, as shown in Fig

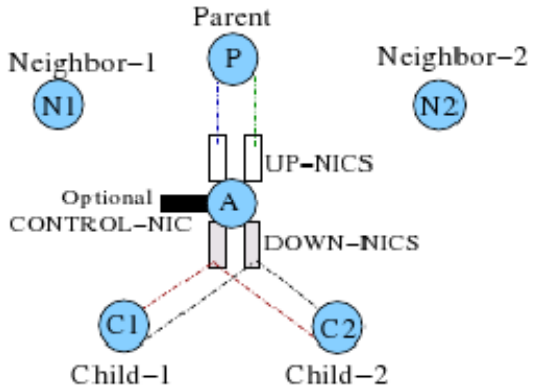


Fig 10 : Eliminating channel dependency problem

Each WMN node is responsible for assigning channels to its DOWN-NICs. Each of the node's UP-NICs is associated with a unique DOWN-NIC of the parent node and is assigned the same channel as the parent's corresponding DOWN-NIC. This restriction effectively prevents channel dependencies from propagating from a node's parent to its children, and thus ensures that a node can assign/modify its DOWN-NICs' channel assignment without introducing ripple effects in the network. Because a gateway node does not have any parent, it uses all its NICs as DOWN-NICs. To increase the relay capability, each non-gateway node attempts to equally divide its NICs into UP- NICs and DOWN-NICs.

Interface-Channel Assignment:

Once the neighbor-to-interface mapping is determined, the final question is how to assign a channel to each of the NICs.

The channel assignment of a WMN node's UP-NICs is the responsibility of its parent.

To assign channels to a WMN node's DOWN-NICs, it needs to estimate the usage status of all the channels within its interference neighborhood. Each node therefore periodically exchanges its individual channel usage information as a CHNL USAGE packet with all its (k + 1)-hop neighbors, where k is the ratio of the interference range and the communication range.

Implementation:

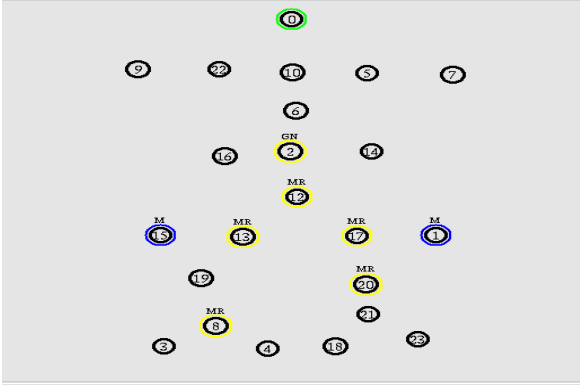


Fig 11: Nodes representation in neighbor interface building .

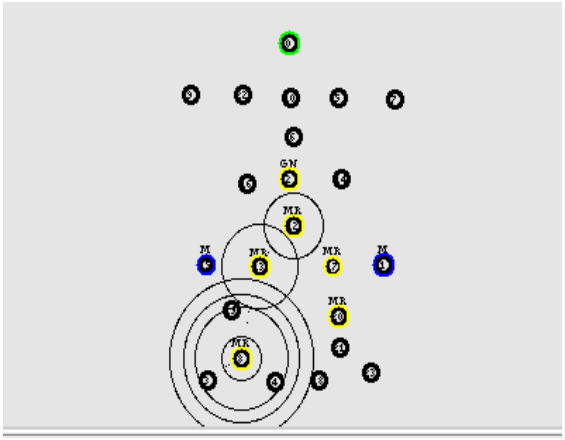


Fig 12: Nodes with shortest path.

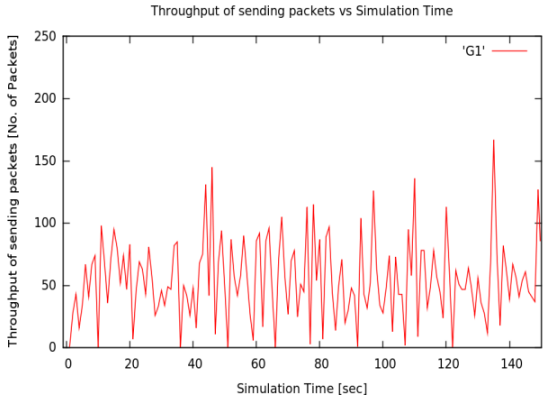


Fig 13: output graph Througput of sending packets vs Simulation Time

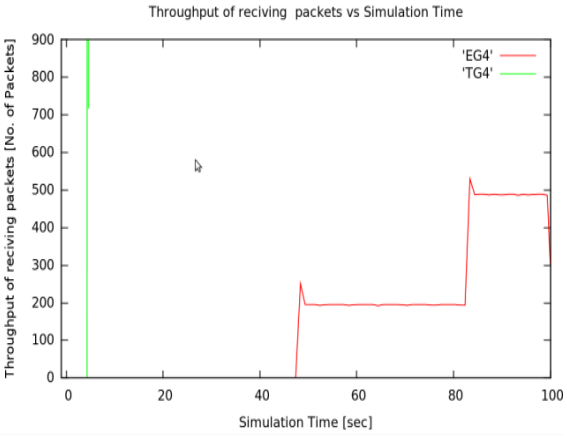


Fig 14: output graph Throughtput of reciving packets vs Simulation Time

REFERENCES :

[1] W. Zhang, G. Xue, J. Tang, and K. Thulasiraman, Faster Algorithms for Constructing Recovery Trees Enhancing QoP and QoS, *IEEE/ACM Transactions on Networking*, Vol. 16 (2008), pp. 642-655.

[2] I. F. Akyildiz, X. Wang, W. Wang, Wireless mesh networks: a survey,*Elsevier Journal of Computer Networks*, Vol. 47, Issue 4, 2005, pp. 445-487.

[3] J. Tang, G. Xue, and W. Zhang, Interference-Aware Topology Control and QoS Routing in

Multi-Channel Wireless Mesh Networks, *ACM MobiHoc'2005*, pp. 68-77.

[4] M. Burkhart, P. von Rickenbach, R. Wattenhofer, A. Zollinger, Does topology control reduce interference, *ACM MobiHoc'2004*, pp. 9-19.

[5] Mohapatra.P.,Wirelessmesh networks,http://www.embeddedwisents.org/dissemination/-pres/padova-mesh-tutorial.pdConner, S.W., IEEE 802.11s Tutorial, Overview of the Amendment for Wireless Local Area Mesh Networking, IEEE802 Plenary, Dallas, Nov, 2006

[6] I. F. Akyildiz, X. Wang, W. Wang, Wireless mesh networks: a survey,*Elsevier Journal of Computer Networks*, Vol. 47, Issue 4, 2005, pp. 445-487.

[7] Asis Nasipuri and Samir R. Das; "A Multichannel CSMA MAC Protocol for Mobile Multihop Networks"; IEEE WCNC 1999.

[8] P. H. Hsiao, A. Hwang, H. T. Kung, and D. Vlah; "Load-Balancing Routing for Wireless Access Networks "; Proc. Of IEEE INFOCOM 2001.

[9] S.J. Lee and M. Gerla; "Dynamic Load-Aware Routing in Adhoc Networks"; Proc. of ICC 2001.

[10] R. Draves, J. Padhye, and B. Zill; "Routing in Multi-radio, Multihop Wireless Mesh Networks" To appear in MobiCom 2004

[11] A. Raniwala, K. Gopalan, T. Chiueh; "Centralized Channel Assignment and Routing Algorithms for Multi-channel Wireless Mesh Networks"; ACM Mobile Computing and Communications Review (MC2R), April 2004.

[12] A. Raniwala, T. Chiueh; "Evaluation of A Wireless EnterpriseBackbone Network Architecture"; To appear in Hot-I 2004.

[13] P. H. Hsiao, A. Hwang, H. T. Kung, and D. Vlah;"Load-Balancing Routing for Wireless Access Networks "; Proc. of IEEE INFOCOM 2001.

◈◈◈

A Two-Step Method for Clustering Mixed Categorical and Numeric Data

B.D. Raveendra Babu, , B.Mahesh Naik

AITS, Rajampeta.
PRRMCE,Shahbad
Email: ravi28888@gmail. Com, mahi_suni2000@yahoo.co.in

Abstract **-** Various clustering algorithms have been developed to group data into clusters in diverse domains. However, these clustering algorithms work effectively either on pure numeric data or on pure categorical data, most of them perform poorly on mixed categorical and numeric data types. In this paper, a new two-step clustering method is presented to find clusters on this kind of data. In this approach the items in categorical attributes are processed to construct the similarity or relationships among them based on the ideas of co-occurrence; then all categorical attributes can be converted into numeric attributes based on these constructed relationships. Finally, since all categorical data are converted into numeric, the existing clustering algorithms can be applied to the dataset without pain. Nevertheless, the existing clustering algorithms suffer from some disadvantages or weakness, the proposed two-step method integrates hierarchical and partitioning clustering algorithm with adding attributes to cluster objects. This method defines the relationships among items, and improves the weaknesses of applying single clustering algorithm. Experimental evidences show that robust results can be achieved by applying this method to cluster mixed numeric and categorical data.

I. INTRODUCTION

With the amazing progress of both computer hard-ware and software, a vast amount of data is generated and collected daily. There is no doubt that data are meaningful only when one can extract the hidden information inside them. However, "the major barrier for obtaining high quality knowledge from data is due to the limitations of the data itself" [1]. These major barriers of collected data come from their growing size and versatile domains. Thus, data mining that is to discover interesting patterns from large amounts of data within limited sources (i.e., computer memory and execution time) has become popular in recent years.

Clustering is considered an important tool for data mining. The goal of data clustering is aimed at dividing the data set into several groups such that objects have a high degree of similarity to each other in the same group and have a high degree of dissimilarity to the ones in different groups[2]. Each formed group is called a cluster. Useful patterns may be extracted by analyzing each cluster. For example, grouping customers with similar characteristics based on their purchasing behaviors in transaction data may find their previously unknown patterns. The extracted information is helpful for decision making in marketing.

Various clustering applications have emerged in diverse domains. However, most of the traditional clustering algorithms are designed to focus either on numeric data or on categorical data. The collected data in real world often contain both numeric and categorical attributes. It is difficult for applying traditional clustering algorithm directly into these kinds of data. Typically, when people need to apply traditional distance-based clustering algorithms (ex., k-means [3]) to group these types of data, a numeric value will be assigned to each category in this attributes. Some categorical values, for example "low", "medium" and "high", can easily be transferred into numeric values. But if categorical attributes contain the values like "red", "white" and "blue" … etc., it cannot be ordered naturally. How to assign numeric value to these kinds of categorical attributes will be a challenge work.

In this paper, a method based on the ideas to explore the relationship among categorical attributes' values is presented. This method defines the similarity among items of categorical attributes based on the idea of co-occurrence. All categorical values will be converted to numeric according to the similarity to make all attributes contain only numeric value. Since all attributes has be-come homogeneous type of value,

existing clustering algorithms can be applied to group these mixed types of data without pain. Nevertheless, most of the existing clustering algorithm may have some limitations or weakness in some way. For example, the returned results from k-means may depend largely on the initial selection of centroid of clusters. Moreover, k-means is sensitive to outliers. In this paper, a two-step method is applied to avoid above weakness. At the first step, HAC (hierarchical agglomerative clustering) [3] algorithm is adopted to cluster the original dataset into some subsets. The formed subsets in this step with adding additional features will be chosen to be the objects to be input to k-means in next step. Since every subset may contain several data points, applying chosen subsets as initial set of clusters in k-means clustering algorithm will be a better solution than selecting individual data. Another benefit of applying this strategy is to reduce the influences of outlier, since the outlier will be smoothed by these added features. The results show that this proposed method is a feasible solution for clustering mixed numeric and categorical data.

The rest of this paper is organized as follows. Next section shows the background and related works. Section 3 describes the proposed method for clustering on mixed categorical and numeric data, and the experimental results will be presented on section 4. Section 5 concludes our work.

II. BACKGROUND

In most clustering algorithms, an object is usually viewed as a point in a multidimensional space. It can be represented as a vector ($x_1...x_d$), a collection of values of selected attributes with d dimensions; and x_i is the value of i-th selected attribute. The value of x_i may be numerical or categorical.

Most pioneers of solving mixed numeric and categorical value for clustering problem is to redefine the distance measure and apply it to existing clustering algorithms. K-prototype [13] is one of the most famous methods. K-prototype inherits the ideas of k-means, it applies Euclidean distance to numeric attributes and a distance function is defined to be added into the measure of the closeness between two objects. Object pairs with different categorical values will enlarge the distance between them. The main shortcomings of k-prototype may fall into followings:

(1) Binary distance is employed for categorical value. If object pairs with the same categorical value, the distance between them is zero; otherwise it will be one. However, it will not properly show the real situation, since categorical values may have some degree of difference. For example, the difference between “high” and “low” shall not equal to the one between “high” and “medium”.

(2) Only one attribute value is chosen to represent whole attribute in cluster center. Therefore, the categorical value with less appearance seldom gets the chance to be shown in cluster center, though these items may play an important role during clustering process. Additionally, since k-prototype inherits the ideas of k-means, it will retain the same weakness of k-means.

Chiu et al. [14] presented a probabilistic model that applies the decrease in log-likelihood function as a result of merging for distance measure. This method improves k-prototype by solving the binary distance problem. Additionally, this algorithm constructs CF-tree [5] to find dense regions to form subsets, and applies hierarchical clustering algorithms on these subsets. Li et al. [15] represents a similarity measure that when two objects have a same categorical value with less appearance in whole data set, greater weight will be assigned to this match. The basic idea is based on Goodall’s similarity measure [16] that the values appearing in uncommon attributes will make greater contributions to the overall similarity among objects. Instead of choosing only one item to represent whole categorical attributes in cluster center, Yin et al. [17] and Ahmad et al. [18] list all items to represent cluster center. The similarity of categorical attributes is calculated based on the proportion of items’ appearance. He et al. [19] calculates the distance between clusters and objects based on all numeric and categorical value’s distribution. The distance is used to decide which cluster an object will belong to.

The major problem of existing clustering algorithms is that most of them treat every attribute as a single entity, and ignore the relationships among them. However, there may be some relationships among attributes. For exam-ple, the person with high incomes may always live in a costly residence, drive luxurious cars, and buy valuable jewelries… and so on. Therefore, in this paper we represent TMCM (a Two-step Method for Clustering Mixed numeric and categorical data) algorithm to solve above problems. The contributions of this proposed algorithm can be summarized as followings:

1. A new idea is presented to convert items in categorical attributes into numeric value based on co-occurrence theory. This method explores the relationships among items to define the similarity between pairs of objects. A reasonable numeric values can be given to categorical items according to the relationship among items.

2. A two-step k-means clustering method with adding features is proposed. K-means’s shortcomings can be improved by applying this proposed method.

In the next session, the TMCM algorithm will be

introduced.

III. TMCM ALGORITHM

In order to explore the relationships among categorical items, the idea of co-occurrence is applied. The basic assumption of co-occurrence is that if two items always show up in one object together, there will be a strong similarity between them. When a pair of categorical items has a higher similarity, they shall be assigned closer numeric values. For example in the instance of Table 1, when temperature is "hot", the humidity is always "high"; but when temperature is "cool", the humidity is "medium" or "low". It indicates that the similarity between "hot" and "high" is higher than the one between "cool" and "high". Therefore, "hot" and "high" shall be assigned a more similar numeric value than "cool" and "high".

The TMCM algorithm is based on above observation to produce pure numeric attributes. The algorithm is shown on Figure 1. Table 2 lists a sample data set, and this data set will be used to illustrate the proposed ideas.

The first step in the proposed approach is to read the input data and normalize the numeric attributes' value into the range of zero and one. The goal of this process is to avoid certain attributes with a large range of values will dominate the results of clustering. Additionally, a categorical attribute *A* with most number of items is selected to be the base attribute, and the items appearing in base attribute are defined as base items. This strategy is to ensure that a non-base item can map to multiple base items. If an attribute with fewer items is selected as the base attribute, the probability of mapping several non-based items to the same based items will be higher. In such a case, it may make different categorical items get the same numeric value.

In step 2 of TMCM algorithm, Attribute X will be selected as the base attribute because it contains the most number of items. Item C, D, and E are defined as base items.

After the based attribute is defined, counting the frequency of co-occurrence among categorical items will be operated in this step. A matrix *M* with n columns and n rows is used to store this information,

Where n is the number of categorical items; m_{ij} represents the co-occurrence between item *i* and item *j* in *M*; m_{ii} represents the appearance of item i.

For example, if a matrix *M* is constructed for the data in Table 2, the value of n will be 5 because there are five categorical items. The value of m_{11} is 4 since item A appears in four transactions in Table 2; and the value of m_{13} is 2 because there are 2 transactions in Table 2 containing both item A and item C. Therefore, the matrix *M* will be:

Table 1. An example of co-occurrence

Temperature	Humidity	Windy
hot	high	false
hot	high	false
cool	low	true
cool	normal	true

TMCM Algorithm

{ // phase 1: Data preprocessing.

1. *Read input data, and normalize numeric attributes.*
2. *Find the attributes A with the most number of items to be base attribute. The items in this base attribute are defined as base items.*
3. *Count the frequency of co-occurrence between every categorical items and every base item, and store these information in a matrix M.*
4. *Using the information in M to build the similarity between every categorical and base item, and store these information in a matrix D.*

// phase 2: Assigning numeric values to categorical items.

5. *Find the numeric attribute that minimizes the within group variance to base attribute. Assign mean of the mapping value in this numeric attribute to every base item.*
6. *Applying the information of similarity that is stored in matrix D to find the numeric values of every categorical item.*

// phase 3: Clustering.

7. *Apply HAC clustering algorithm to group data set into i clusters. (In this paper, i is set to the 1/3 of number of objects.)*
8. *Calculate the centroid of every formed cluster, and add every categorical item to be additional attributes of centroid. The value of a new attribute is the number that objects in this cluster contains this item.*
9. *Applying k-means clustering algorithm again to group formed clusters in step7-step8 into desired k groups.*

Figure 1. The TMCM algorithm.

Table 2. A sample data set

Attribute W	Attribute X	Attribute Y	Attribute Z
A	C	0.1	0.1
A	C	0.3	0.9
A	D	0.8	0.8
B	D	0.9	0.2
B	C	0.2	0.8
B	E	0.6	0.9
A	D	0.7	0.1

$$M = \begin{pmatrix} 4 & 0 & 2 & 2 & 0 \\ 0 & 3 & 1 & 1 & 1 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Since the frequencies of co-occurrence between base items and other categorical items is available by retrieving the elements in matrix *M*, the similarity between them can be calculated by adopting following equation:

$$D_{xy} = \frac{|m(X,Y)|}{|m(X)| + |m(Y)| - |m(X,Y)|} \quad (1)$$

where X represents the event that item x appears in the set of objects; Y represents the event that item y appears in the set of objects; m(X) is the set of objects containing item x; m(X,Y) is the set of objects containing both item x and y.

In equation (1), when two items always show up together in objects, the similarity between them will be one. If two items never appear together, it will get zero for the similarity measure. The higher value of D_{xy} means the more similar between item x and item y. However, only the values of D_{xy} larger than a threshold will be recorded, or zero will be assigned. Now the similarity between every categorical item and every base item is available. For example, the value of |m(A)| is 4 which can be obtained from m_{11} in matrix *M*. Similarly, The value of |m(A, C)| is 2 because m_{13} in matrix *M* is 2. Therefore,

D_{AC} = 2 / (4 + 3 Ę 2) = 0.4,

D_{AD} = 2 / (4 + 3 Ę 2) = 0.4,

D_{AE} = 0/ (4 + 1 Ę 0) = 0.

The first process in phase 2 is to find the numeric attribute that minimizes the within group variance to base attribute. The equation for within group variance will be

$$SS_w = \sum_j \sum_i (X_{ij} - \overline{X_j})^2 \quad (2)$$

Where $\overline{X_j}$ is the mean of mapping numeric attribute of *j-th* base item.

X_{ij} is the value of *i-th* value in mapping numeric attribute of *j-th* base item.

Attributes Y in Table 2 is identified to meet this requirement. Then, every base item can be quantified by assigning mean of the mapping value in the selected numeric attribute. For example, the value of item C in Table 2 is (0.1 + 0.3 + 0.2) / 3 = 0.2, item D is 0.7 and item E is 0.6.

Since every base item has been given a numeric value, all other categorical items can be quantified by applying the following function.

$$F(x) = \sum_{i=1}^{d} a_i * v_i \quad (3)$$

Where *d* is the number of base item; a_i is the similarity between item *x* and i-th base item; v_i is the quantified value of *i-th* base item.

Therefore, item A in Table 2 will be assigned the following value: F(A) = 0.44 * 0.2 + 0.44 * 0.7 + 0 * 0.6 = 0.396.

All attributes in data set will contain only numeric value at this moment, the existing distance based clustering algorithms can by applied without pain. HAC (Hierarchical Agglomerative Clustering) is a widely used hierarchical clustering algorithm. Several HAC algorithms have appeared in the research community. The major difference is the applied similarity criteria. The HAC algorithm takes numeric data as the input and generates the hierarchical partitions as the output. Therefore it is applied in first clustering step to group data into subsets. In HAC, initially each object is considered as a cluster. Then by merging the closest clusters iteratively until the termination condition is reached, or the whole hierarchy is generated. It generates different levels of clusters bottom-up. The algorithm of HAC is presented in Figure 2.

The k-means algorithm takes numeric data as input, and generates crispy partitions (i.e., every object only belongs to one cluster) as the output. It is one of the most popularly used clustering algorithms in the research community. It has been shown to be a robust clustering method in practice. Therefore, the k-means algorithm is applied in second clustering step to cluster data sets. K-means starts by randomly selecting or by specifically picking k objects as the centroids of *k* clusters. Then k-means iteratively assigns the objects to the closest centroid based on the distance measure, and updates the mean of objects in this cluster as the new centroid until reaching a stopping criterion. This stopping criterion could be either non-changing clusters or a predefined number of iterations. The algorithm of HAC is presented in Figure 3.

Because k-means suffers from its shortcomings mentioned in previous section, a two-step method is introduced to improve it. Initially, this proposed method applies HAC to group data set into *i* subsets, where *i* is set to the one-third of number of objects in this paper. Based on the observations of clustering results, these settings of *i* yield satisfied solutions. Each formed subsets will be treated as an input object for applying k-means in next step. The centroid of each subset is

calculated to represent whole subset. Moreover, every categorical item will be added to be additional attributes of centroid. The value of a new attribute is the number that objects in this cluster contains this item.

1. *Calculate the distance between every two objects.*
2. *View each object as an individual cluster.*
3. *Merge the closest two clusters*
4. *Update the distance between clusters.*
5. *Repeat 3-4 until reaching a stopping criterion or generating the whole hierarchy.*

Figure 2. HAC algorithm.

1. *Select first k objects randomly as the centroid of each cluster.*
2. *Assign each object to the closest cluster based on Euclidean distance or cosine similarity.*
3. *Update the centroid of each cluster.*
4. *Repeat steps 2-3 until stopping criterion is reached.*

Figure 3. K-means algorithm.

V. CONCLUSION

Clustering has been widely applied to various domains to explore the hidden and useful patterns inside data. Because the most collected data in real world contain both categorical and numeric attributes, the traditional clustering algorithm cannot handle this kind of data effectively. Therefore, in this paper we propose a new approach to explore the relationships among categorical items and convert them into numeric values. Then, the existing distance based clustering algorithms can be employed to group mix types of data. Moreover, in order to overcome the weaknesses of k-means clustering algorithm, a two-step method integrating hierarchical and partitioning clustering algorithms is introduced. The experimental results show that the proposed approach can achieve a high quality of clustering results. In this paper, the TMCM algorithm integrates HAC and k-means clustering algorithms to cluster mixed type of data. Applying other algorithms or sophisticated similarity measures into TMCM may yield better results.

Furthermore, the number of subset *i* is set to one-third of number of objects in this paper. Although experimental results show that it is feasible, how to set this parameter precisely is worth more study in the future.

REFERENCES

[1] Wiederhold, G., Foreword. In: Fayyad U., Shapiro G.P., Smyth P., Uthurusamy R., editors, Advances in Knowledge Discovery in Databases. California: AAAI/MIT Press, 1996;2.

[2] Han, J. and Kamber, K., Data mining: Concept andTechniques. San Francisco: Morgan Kaufman Publisher (2001).

[3] Jain, A. K. and Dubes, R. C., *Algorithms for Clustering Data*, New Jersey: Printice Hall (1988).

[4] Kaufman, L. and Rousseeuw, P. J., *Finding Groups in Data: An Introduction to Cluster Analysis*, New York: JohnWiley & Sons (1990).

[5] Ng, R. and Han, J., *Efficient and Effective ClusteringMethod for Spatial Data Mining*, Proc of the 20thVLDB Conf. 1994 September. Santiago, Chile (1994).

[6] Zhang, T., Ramakrishman, R. and Livny, M., *BIRCH:an Efficient Data Clustering Method for Very LargeDatabases*, Proc. 1996 ACM-SIGMOD Int. Conf. Management of Data, 1996 June. Montreal, Canada (1996).

[7] [7] Guha, S., Rastogi, R. and Shim, K., *Cure: An Efficient Clustering Algorithm for Large Databases*, Proc. 1998ACM-SIGMOD Int. Conf. Management of Data. 1998 june. Seattle,WA(1998).

[8] Ester, M., Kriegel, H. P., Sander, J. and Xu, X., *A Density-Based Algorithm for Discovering Clusters inLarge spatial databases*, Proc. of the Second International Conference on Data Mining (KDD-96), 1996 August. Portland, Oregon (1996).

[9] Hinneburg, A. and Keim, D., *An Efficient Approach to Clustering in Large Multimedia Databases with Noise*, Proc. 1998 Int. Conf. on Data Mining and Knowledge Discovery (KDD'98). 1998 August. New York (1998).

[10] Wang, W., Yang, J. and Muntz, R., *Sting: A Statistical Information Grid Approach to Spatial Data Mining*, Proc. 23rd VLDB. 1997 August. Athens,Greece (1997).

[11] Kaufman, L. and Rousseeuw, P. J., *Finding Groups in Data: An Introduction to Cluster Analysis*, New York: JohnWiley & Sons (1990).

[12] Hinneburg and Keim, D., *An Efficient Approach to Clustering in Large Multimedia Databases with Noise*, Proc. 1998 Int. Conf. on Data Mining and Knowledge Discovery (KDD'98). 1998 August. New York (1998).

[13] Huang, Z., "Extensions to the K-Means Algorithm for Clustering Large Data Sets with Categorical Values," *Data Mining and Knowledge Discovery*, Vol. 2, pp.283□304 (1998).

[14] Chiu, T., Fang, D., Chen, J. and Wang, Y., *A Robust and Scalable Clustering Algorithm for Mixed Type Attributesin Large Database Environment*, Proc. 2001 Int. Conf. On knowledge Discovery and Data Mining. 2001 Auguest. San Fransico (2001).

[15] Li, C. and Biswas, G., "Unsupervised Learning with Mixed Numeric and Nominal Data," *IEEE Transactions on Knowledge and Data Engineering*, Vol . 14, p.4 (2002).*18 Ming-Yi Shih et al.*

[16] Goodall, D. W., "A New Similarity Index Based on Probability," *Biometric*, Vol. 22, pp. 882□907 (1966)

◈◈◈