IITM Journal of Information Technology

Volume 3 (January - December 2017)
ISSN No. 2395 – 5457
SPECIAL ISSUE

IITM
Nurturing Excellence

INSTITUTE OF INNOVATION IN TECHNOLOGY & MANAGEMENT
Affiliated to GGSIP University, 'A' Grade by GNCTD & Assessed as 'A+' by SFRC
NAAC Accredited Institute
(ISO 9001-2008 Certified Institute)

NCIRCSIT'2017
National Conference on Innovative Research in Computer Science and Information Technology
February 18, 2017

IN TECHNICAL COLLABORATION WITH
COMPUTER SOCIETY OF INDIA (CSI)

SPONSORED BY
DEFENCE RESEARCH AND DEVELOPMENT ORGANIZATION (DRDO)

Editors: Prof. (Dr.) Geetali Banerji, Prof. Ankit Verma

IITM Journal of Information Technology

IITM Journal of Information Technology is a National Annual Journal of Information Technology intended for professionals and researchers in all fields of Information Technology. Its objective is to disseminate experiences, ideas, case studies of professionals in Information Technology to propagate better understandings. Its focus is on empirical and applied research and reflections that are relevant to professionals of Information Technology with academic standards and rigor within purview. The views expressed in the Journal are those of authors. No part of this publication may be reproduced in any form without the written consent of the publisher. The Journal intends to bring together leading researchers and Information Technology practitioners from around the country.

IITM Journal of Information Technology

Annual Journal of Institute of Innovation in Technology & Management

Volume 3 **January-December, 2017**

CONTENTS

Research Papers & Articles

BLANK

Swarm Intelligence Based Routing Protocols in Wireless Sensor Networks

Vishakha Sehdev*
Ankit Verma**

Abstract

Wireless sensor networks are made up of small sensors which are useful in sensing and transmitting data. It is very popular field of computer science as sensors are very small, cheap and optimize communication devices. Swarm intelligence is novel field which is study of insects and based upon behavior of insects. Several protocols and algorithms are defined using swarm intelligence to optimize performance of wireless sensor networks.

Key Words: swarm intelligence, wireless sensor networks, performance parameters, routing protocols.

I. Introduction

Wireless Sensor networks sometimes known as wireless network consists of many small scale sensors that communicate via wireless channel and is used to monitor Physical or environmental conditions like sound, pressure, temperature . A WSN typically consist of a large number of sensor nodes which have low-cost, low-power and are multi-functional. These nodes have certain restrictions like limited energy, bandwidth, transmission range and storage because of these restrictions MANETs and end to end devices are used in many routing schemes. Wireless sensor nodes have small size, these nodes are not only capable of sensing the data but also have capability of processing and communicating the data. Cost, connection errors and deployment delay can be reduced using wireless sensor network. They major goal is to extend the network life during searching the shortest paths from the source nodes to the base node [1]. Task allocator, base station and sensor nodes are major parts of Sensor network. The sensor network consists of region which is under surveillance. The exact position of sensors is impractical to find because sensor units are randomly scattered to the region. These sensors units have minute size, low power utilization and are capable of short range communication. Figure 1 shows components of wireless sensor unit.

Vishakha Sehdev*
Department of Computer Science & Engineering
GITM, Gurgaon, Haryana

Ankit Verma**
Department of Information Technology
IITM, Janakpuri, New Delhi

II. Routing Protocol Design Factors

Deployment of WSN was first done by Military applications, and later internet which is an instance of civilian application was pursued. The working of routing protocols is different than the working of wired network such as internet because of the structural difficulties of WSN.

- *Quality of Service:* Specific applications require quality of service (QoS) which can be thought of as service excellence. The quality of service can be measured in terms of length of life span, data reliability, location-awareness, energy efficiency and collaborative-processing, and it is dependent of application domain which influences the selection of routing protocols for that application.
- *Energy:* Sensor nodes have limited amount of energy. This limited energy drain while performing task such as transmitting and information computations in a wireless environment.

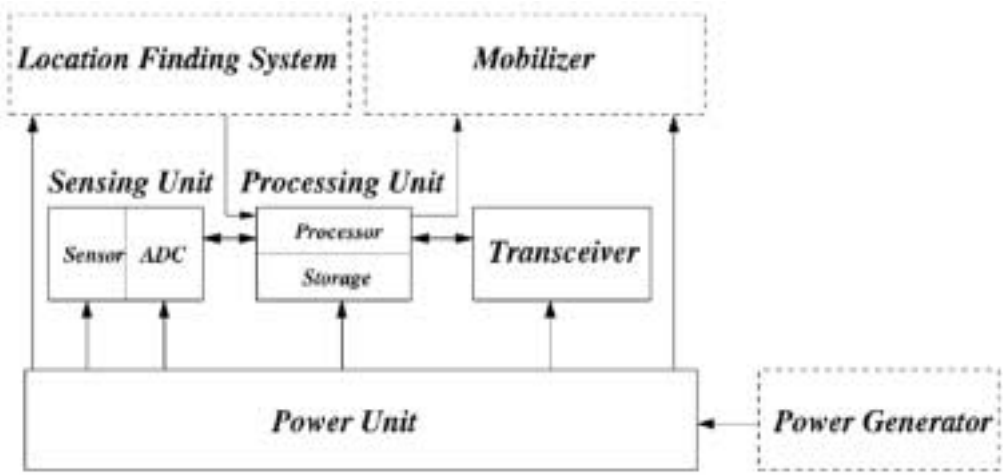


Figure 1: Components of Wireless Sensor Unit

- *Operating Environment:* Sensor nodes can be deployed to woods, oceans and inside of a building or different areas and it depends on necessities of the applications [4]. Routing protocols is highly affected by changing environment.
- *Scalability:* Sensor applications may have numerous of sensor nodes and because of it scalability is a considerable property of a routing protocol. The number of sensor nodes deployed in the sensing area may vary from hundreds, thousands or more and therefore routing algorithms must be scalable enough to react to events.
- *Data delivery models:* Data are flooded from the sensor nodes to the sink or base station in a sensor node. Every node transmits the data periodically in the continuous model [3]. Events are run in a triggered manner in Event driven model. The data on query of the sink is initiated by the Query driven model.
- *Network life span:* Network Life Span depends on various applications and it may vary from hours to several years. Overall network mission shouldn't be affected by the failure of sensor nodes such as physical damage, blockage due to lack of power, or environmental interference.
- *Data assembly:* Huge traffic over wireless sensor network can be generated from transmission of trivial data. For avoidance of traffic, similar packets from several nodes can be merged.
- *Latency:* In WSN, Latency can be thought of as time taken for a packet of data to travel from sensor point to the base station, latency is measured either one-way (the time taken for the source to send a packet to the destination receiving it) or round-trip (one-way latency from source to destination, including the one-way latency from the destination back to the source).

III. Swarm Intelligence Based Routing Protocols For Wireless Sensor Networks

Typical sensor network operates in five phases: the planning phase, deployment phase, post-deployment phase, operation phase and post-operation phase [5]. There are three categories of Swarm based routing protocol: Bee based, Ant based, and slime based.

A. Bee Colony Based Routing Protocol

Honeybees foraging behaviors is the motivation of Bee Colony based routing protocol. Mechanisms for WSNs such as self-organization and division of labor are provided by honeybees. Inspired from bees, there are a few routing protocols for WSNs

- ***Bee Sensor:*** The bee-hive routing protocol was first developed for wired network but later on it was implemented on wireless sensor networks. Bee-sensor algorithm is based on the foraging principles of honey bees with an on-demand route discovery (AODV). The three types of agents used by this

approach are scouts, packers and foragers bees [11]. At source node, Packers are responsible for locating Packers appropriate foragers for the data packets whereas scouts responsibility of scouts is discovering the path to a new destination. Foragers are responsible for carrying the data packets to a sink node.In this approach, interaction is done between scouts and source routing and based on this small forwarding table are made during the return of scout.

B. Slime Mold Based Routing Protocol

In Slime mold based routing protocol, the term slime mold is basically used to the heterotrophic organism like unicellular and fugues. Such unicellular organisms have strong resemblance between them. Wireless sensor network consist of nodes which are designed to perform simple task, these nodes have capacity to react and have characteristics like limited capacity and simplicity.

- ***Multi-sink swarm-based routing protocol:*** The characteristics of WSN include fault tolerance, environmental adaption and self organization. Slime mold organisms are the inspiration behind these protocols. By the use of pheromone generation and evaporation functions, slime mold organism organize themselves in clusters. With the help of gradient concept this direct data traffic towards multiple sink nodes and shows fault tolerance and autonomy.

C. Ant Colony Optimization

Foraging behavior of ant species is the major inspiration of Ant Colony Optimization. ACO use the mechanism of ant species for solving optimization problems. These ant species are responsible for producing Favorable paths which is followed by the others ant of the colony by depositing pheromone on the ground. Every time when transfer of data from source node to destination node is done the ants are directed from source node to the destination node. An experiment named "double bridge experiment was performed on the ant species where two bridge of equal length provided connection between food source and nest of the colony [7]. During the experiment, the ants started discovering the environment and reach to their destination it means they arrive at the food source. Pheromone was deposit by the ants while discovering their path between nest and food source. Random path was chosen by the ants initially but after some time it was noticed that out of the two bridges the pheromone concentration was higher on one bridge [10]. This higher concentration of pheromone attracted more ants of the colony and after some time the use of that bridge started by the whole colony of ants.

- ***Energy-efficient ant-based routing (EEABR):*** EEABR uses data structure which is responsible for storing ant information. A router table is used which stores the forward node, previous node, the ant identification and a timeout value. The node looks at the routing table if node receive forward ant and start searching the ant identification for a loop [12]. The node stores the information, restart the timer and forward that ant to the next node if ant identification is not found but if an ant identification is found, that ant is being eliminated. If a backward ant is received, the next node is finding by searching the routing table.The record which identifies the backward ant is deleted by the Timer if ant doesn't reach that node within the time defined by the timer.
- ***Sensor-driven and cost-aware ant routing (SC):*** In Sensor driven and cost ware ant routing, the sensors sense the finest direction in which ants go initially which in turns increase the performance of the forward ants. Each node is accountable of saving the cost to the destination from each of the neighbors and also stores the probability distribution.
- ***Flooded forward ant routing (FF):*** When there is a case where initially destination is undefined and cost estimation cannot be done the ants in the colony starts investigating the area. Flooding of ants ot the route packets destination is the approach behind Flooded forward ant routing. In this routing, broadcast functionality is used.
- ***Flooded piggybacked ant routing (FP):*** In this approach data ants are used which carrys the forward

list which is used to forward ants. . In contrast, control of the flooded forward ants are treated same as in FF for flooded data ants. Flooding is a dynamic, distributed and complex task.

- ***Ant colony optimization-based location-aware routing (ACLR):*** The idea behind this algorithm is that the ants instead of whole neighbors which assures the data packets delivery selects the next hop from the subset of set of node neighbors. A formula is used to measure the transition probability which helps ants in selecting the next hop. A model is used to determine the amount of pheromone deposited by the ants.

IV. Conclusion

Because of the dynamic and distributed behavior of networks there is a requirement of new routing approaches. Limitations such as adaptability, scalability, battery life, and cost and so on can be removed by the use of swarm intelligence based routing protocols. In comparison to other approaches ant based optimization is gaining attention by a lot of researchers.

References

1. Heinzelman W B, Chandrakasan A P, Balakrishnan H. An application-speciûc protocol architecture for wireless micro- sensor networks. IEEE Trans Wirel Commun, 2002, 1: 660–670.
2. Ben-Othman J, Yahya B. Energy eûcient and QoS based routing protocol for wireless sensor networks. J Parallel Distrib Comput, 2010, 70: 849–857.
3. Lin F Y S, Yen H H, Lin S P. Delay QoS and MAC aware energy-eûcient data-aggregation routing in wireless sensor networks. Sensors, 2009, 9: 7711–7732.
4. Wang H P, Zhang X B, Na¨ýt-Abdesselam F, et al. Cross-layer optimized MAC to support multihop QoS routing for wireless sensor networks. IEEE Trans Veh Technol, 2010, 59: 2556–2563.
5. R. S. Parpinelli, H. S. Lopes, and A. A. Freitas, "Data mining with an ant colony optimization algorithm," IEEE Transactions on Evolutionary Computation, vol. 6, no. 4, pp. 321–332, 2002.
6. K. M. Sim and W. H. Sun, "Ant colony optimization for routing and load-balancing: Survey and new directions," IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 2003.
7. V. A. Cicirello and S. F. Smith, "Ant colony control for autonomous decentralized shop ûoor routing," in Proceedings of the International Symposium on Autonomous Decentralized Systems. IEEE Computer Society Press, 2001, pp. 383–390.
8. Kamat, P.; Zhang, Y.; Trappe, W.; Ozturk, C. Enhancing Source-Location Privacy in Sensor Network Routing. In Proceedings of the 25th IEEE Int. Conference on Distributed Computing Systems (ICDCS05), 2005; pp. 599-608.
9. Al-Karaki, J.N.; Kamal, A.E. Routing Techniques in Wireless Sensor Networks: a Survey. IEEE Wirel. Commun. 2004, 11, 6-28.
10. Iyengar, S.; Wu, H.-C.; Balakrishnan, N.; Chang, S.Y. Biologically Inspired Cooperative Routing for Wireless Mobile Sensor Networks. IEEE Syst. J. 2007, 1, 29–37.
11. Niannian D.; Liu, P.X.; Chao H. Data Gathering Communication in Wireless Sensor Networks Using Ant Colony Optimization. In IEEE/RSJ International Conference on Intelligent Robots and Systems, 2005; pp. 697–702.
12. Al-Karaki JN, Kamal AE (2004). "Routing Techniques in Wireless Sensor Networks: a Survey". Wireless Communications, IEEE, 11: 628.
13. Ding N, Liu XP (2005). A Centralized Approach to Energy-Efficient Protocols for Wireless Sensor Networks, IEEE International Conference on Mechatronics and Automation, Niagara Falls, Ont., Canada, 3: 1636-1641.

Demonetization: A Blessing in Disguise for IT Industry in India

Sonal Anand*
Lubna Ansari**

Abstract

The paper highlights the impact of demonetization on e-commerce in India, and how Indian IT industry has benefited from it. Further, we discuss about a wide range of IT applications which work on cashless payments, and the collective role of government, citizens and IT sector to transform India into a cashless society.

Key Words: cashless transactions, demonetization, digital payment, e-commerce, e-wallet

I. Introduction

Demonetization refers to depriving a currency unit of its status, .i.e. withdrawal of a currency unit from circulation. With effect from 9th Nov 2016, demonetization of Rs 500 and Rs 1000 notes was announced by Mr. Narendera Modi, the Prime Minister of India, which led to a noticeable impact on Indian e-commerce, and majorly on IT applications that focus on cashless transactions.

E-commerce refers to the commercial transactions that are conducted electronically, where transmitting of funds is carried out primarily over Internet. The purpose of demonetization has been stated as "moving towards a cashless economy and making India a cashless society". The public has been encouraged to opt for cashless transactions.

A variety of IT applications have benefited from the decision, one of which is the very well known, Paytm, which provides services like online recharge, bill payment, DTH recharge, online transactions and others. The success of Paytm is an inspiration for IT startups in India and the commencement of Prime Minister's "Make in India" initiative. According to our survey conducted among a limited set of urban population in New Delhi, while a major section of people are in support of digital payments, a large percentage of people find it impractical, inconvenient and an insecure mode of transaction. Lastly, we discuss about the future scope of IT startups and the role of government and IT sector to make the decision practical, feasible and acceptable among masses.

Sonal Anand*
Assistant Professor (IT)
IINTM, Janakpuri, GGSIPU
Lubna Ansari**
Assistant Professor (IT)
IINTM, Janakpuri, GGSIPU

II. Literature Review

Quartz India has discussed how demonetization can lead to a new digital economy in India. As per an Assocham-RNCOS joint study, the mobile payment transactions are expected to register a growth rate from 3 million transactions to 153 billion by year 2021-22[1]. Karan Kashyap from Forbes has discussed about the increase in the number of daily transactions through e-wallet, since the announcement of demonetization[2]. According to Business Insider , Paytm has benefited the most, among all e-commerce companies in India so far. Paytm simultaneously simplified the method of using the app for the new users [3]. With over 850,000 offline merchants using Paytm across India, Paytm is also looking to target daily wage workers such as electricians or plumbers .[4] An article in Forbes has discussed the positive aspects of demonetization, and how, despite the widespread inconvenience and resistance, 1.25 billion

Indians has supported the decision.[5]. As per Daily News and Analysis, A major group of citizens including politicians, businessmen and others are against the decision, citing the need for transparency and release of a white paper demanding the gain and loss incurred by the decision[6]. India Arising has explained numerous benefits of demonetization and how it is going to revolutionize the nation[7]. It has helped in eliminating counterfeit currency from circulation, played a major role in curbing terrorism, advertised the use of mobile ATMs, eradicated stone pelting and riots in terror prone areas in Hawala, led to solving 55 lakh monetary cases in Lok Adalats in one day, made people deposit more than 55 lakh crore money in banks in just first 15 days, deposition of money in Jan Dhan accounts, has made businessmen deposit crores in the name of advance tax, compelled defaulters to pay their pending taxes, loans and electricity bills, and provided Income Tax departments with helpful information about suspected defaulters. Several small scale vendors have started going digital and are already using e-wallet and net banking services for payments. The unaccounted money is definitely going to extinguished, and it is eventually going to add to the economic growth in India. The purpose is also to implement a habit changing effect among Indians, and to make them deposit funds in banks rather than keeping them at homes. With lower cash transactions, the inflation definitely seems to reduce.

III. Impact on IT

As per Forbes, the number of daily transactions through e-wallet services such as Paytm, FreeCharge and Mobikwik has increased from 1.7 million — recorded on November 8th 2016 to 6.3 Million as of December 7th, 2016. Mobikwik has gained 5 million new users since the announcement.

Quartz India talks about India's first "digital and cashless village," Akodara which has been highlighted after the news, where mobile banking has been active for years and the villagers have been making cashless payments for buying everything from grocery to bill payment, way before demonetization was announced.

This is a major inspiration for a country where a majority of people use cash for transactions, and banking is still considered as an unachievable luxury by people below poverty line.

Currently Paytm is witnessing 7 million transactions worth Rs 120 crore in a day. The overnight success of Paytm is an opening gateway for budding IT companies. The sudden need for cashless payment makes way for mobile applications catering a variety of service areas. A few of these include grophers for online food shopping, amazon, jabong, flyrobe, abof, myntra for apparel purchase, Mobikwik, FreeCharge, and SBI buddy for banking and commercial transactions, Ola money for cab payments, IRCTC for online travel booking, MakeMyTrip for deals on travels and hotels, zomato for checking for and reviewing restaurants, nearbuy for location based booking of services, urban clap to book at-home services, and a lot more.

IV. Survey Report

As per our survey conducted among a group of teenagers, and adults among different age categories in urban areas, we found miffed responses by people in support of or against digital currency.

A. Reasons for acceptance of digital currency

- **Convenience**

Cashless transactions are reportedly more convenient than cash transactions because of considerable amount of saving in time and effort of visiting banks.

- **Easy track of finances**

Spending cash doesn't help customers keep account of their monthly or yearly expenditures and balances, while net banking and other electronic modes of payments help customers keep track of finances, which can then be more effectively managed.

- **Rewards**

E-wallet and other electronic services are banking upon their promotion codes, cash back and other schemes to attract customers, saving them ample amount of cash in the long run.

- **Fear of future demonetization**

It's better to opt for exchange of digital currency in comparison to cash, because of a chance of a future

demonetization on currency notes of other denomination.

B. Reasons for non acceptance of digital currency

- **Resistance towards change**

In a country where majority of payments are made with cash, it's practically impossible to expect people to change their payment modes overnight.

- **Illiteracy**

The absence of awareness about banking, e-wallets, digital currency and other online services among millions of citizens is a major roadblock in the fulfillment of the purpose.

- **Poverty**

In a country where a major section cannot even afford food, shelter and basic amenities of life, expecting people to have smart phones and use mobile applications is still a distant dream.

C. Security threats in electronic transactions

For the literate, aware and educated percentage of the population who can afford and easily adopt the cashless mode of payment, the security threats in electronic transactions are major obstacles, which hinder their interests towards adopting this mode of payment.

V. Future Scope of IT Start-ups

As India moves from being a pro-cash to a cashless society, the role of IT seems to expand hereon, making way for budding IT startups which work on digitizing services in exchange of electronic currency. After the overnight success of Paytm, there is a surety of increase in demand for IT apps catering to daily services like travelling via e-rickshaws which accept digital currency and ordering daily products like medicines online. Another way of digitizing the economy can be issuing a common smartcard for every possible transaction like buying articles, as well as for paying travel expenses.

Demonetization has also made way for USSD, i.e Unstructured Supplementary Service Data, which is a technology by GSM (Global System for Mobile) which uses prepaid roaming or mobile chatting to connect application programs with mobile devices. USSD makes use of WAP (Wireless Application Protocol) to provide location based services, mobile banking and call back services, by establishing a real time, two way communications for exchange of data, and thereby eliminating the need of an internet connection for electronic transactions. [8]

As per our survey, the urban youth is majorly in support of digital payments, and expects an enhancement in services provided by IT sector, from smart cards to pay their bus fares, to buy medicines online, to pay for their tuition and college fees through e-wallet, to paying hospital bills, water and electricity bills and a lot more.

Like there are two sides of a coin, adapting to a cashless society is not easy, for millions of Indians, due to unreachablility, restrict to change, illiteracy and security issues in electronic transactions

The enlightening of the poor, the old and the illiterate is an important task. A lot needs to be worked upon, and it's the collective responsibility of Indian government and IT sector to work towards a secure e-commerce environment, which can be easily be adapted by people.

VI. Conclusion

Demonetization may have brought a temporary slowdown in Indian economy, but with a collective effort, it can promise a new digital economy with a great scope for Indian IT startups which work on cashless transactions. Digital payments may be the need of the hour, but with a focused effort, the Prime Minister's dream of a cashless society can come true.

References

1. Dwivedi Y., *http://qz.com/843872/indias-rupee-demonetization-could-spark-a-new-digital-economy-in-the-cash-reliant-country/#*, Quartz India,2016
2. Kashyap,K.,*http://www.forbes.com/sites/krnkashyap/2016/12/15/how-indias-demonetization-is-affecting-its-startups/*,Forbes 2016

Data Mining Techniques and Applications in Cloud Computing

Dr. Rashmi Jha*

Abstract

In cloud computing paradigm data mining techniques and applications are very much needed. Data mining techniques through Cloud computing will allow the users to retrieve meaningful information from virtually integrated data warehouse. Beside this, it also reduces the costs of infrastructure and storage. This paper describes how the Cloud Computing has become a main source for the data processing, storage and distribution. Data mining has been an effective tool to analyze data from different perspective and getting useful information from data, classification of data, categorization of data, and to find correlation of data patterns from the dataset. Into normal day-to-day activities the integration of data mining techniques has become common place. Data mining techniques help businesses to become more efficient by reducing costs.

Key Words: Cloud computing, data mining, data mining in cloud computing.

I. Introduction

Large data set has become a major challenge and data intensive computing is now considered as the "fourth paradigm" in Scientific discovery after theoretical, experimental, andcomputation science. Data mining has been an effective tool to analyze data from different angles and getting useful information from data.

The internet is becoming an increasingly vital tool in everybody's life, both professional and personal, as its user and becoming more numerous. The most revolutionary concept of recent year is Cloud Computing. Many companies are choosing as an alternative to building their own IT infrastructure to host database or software, having a third party to host them on its large servers, so company's would have access to its data and software over the Internet.

The use of cloud computing is gaining popularity due to its mobility, huge availability and low cost. On the other hand it brings more threats to the security of the company's data and information. In recent years, data mining techniques have evolved and become more used, discovering knowledge in database becoming increasingly vital in various fields: business, medicine, science and engineering, spatial data etc.

Dr. Rashmi Jha*
Institute of Innovation in Technology and Management
New Delhi

II. Important Aspects of Cloud Computing

Cloud computing represents both the software and the hardware delivered as services over the Internet. Cloud Computing is a new concept that defines the use of computing as a utility, that has recently attracted significant attention.

- **Phase 1:** people used terminals to connect to powerful mainframes shared by many users.
- **Phase 2:** stand-alone personal computers became powerful enough to satisfy users' daily work.
- **Phase 3:** computer networks allowed multiple computers to connect to each other.
- **Phase 4:** local networks could connect to other local networks to establish a more global network.
- **Phase 5:** the electronic grid facilitated shared computing power and storage resources.
- **Phase 6:** Cloud Computing allows the exploitation of all available resources on the Internet in a scalable and simple way.

Table I. Top Computing Companies and Key Features

Cloud Name	Key Feature
Sun Microsystems Sun Cloud	Sun Cloud More available application than any other open os
IBM Dynamic Infrastructure	Integrated power management to help you plan, predict, monitor and activity manage power actively manage power consumption of your Blade Centre servers.
Amazon EC2	Designed to make wed-scale computing easier for developers
Google App Engine	No limit to free trial period if you do not exceed the quota allotted.
AT&T Synaptic Hosting	Use fully on-demand infrastructure or combine it with dedicated components to meet specialized requirements.
Go Grid Cloud computing	Free load balancing and free 24/7 supports
Salesforce	Offers cloud solutions for automation and platform system performance and security at trust.salesforce.com
Microsoft Azure	Currently offering a "development accelerator" discount plan. 15-30 % discount off consumption changes for first 6 months. AT

III. Data Mining Techniques

Data mining is defined as a "type of database analysis that attempts to discover useful patterns or relationships in a group of data. The analysis uses advanced statistical methods, such as cluster analysis, and sometimes employs artificial intelligence or neural network techniques. A major goal of data mining is to discover previously unknown relationships among the data, especially when the data come from different databases." The most important data mining techniques are-

Table II. Data Mining Techniques

Cloud Name	Key Feature
❖ Clustering	Useful for exploring data and finding natural groupings. Members of a cluster are more like each other than they are like members of a different cluster. Common examples include finding new customer segments and lifesciences discovery
❖ Classification	Most commonly used technique for predicting a specific outcome such as response/ noresponse, high / medium / low value customer, likely to buy / not buy.
❖ Association	Find rules associated with frequently occurring items, used for market basket analysis, cross-sell, root cause analysis. Useful for product bundling, instore placement, and defect analysis. Regression
❖ Regression	Technique for predicting a continuous numerical outcome such a customer process yield rates.
❖ Attribute Importance	Ranks attributes according to strength of relationship with target attribute. Use cases include finding factors most associated with customers who respond to an offer
❖ Anomaly Detection	Identifies unusual or suspicious cases based on deviation from the norm. Common examples include health care fraud, expense report fraud, and tax compliance
❖ Feature Extraction	Produces new attributes as linear combination of existing attributes. Applicable for text data decomposition and projection and pattern recognition

Data mining is the extraction of hidden predictive information from large database, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouse. Data mining tools predict future trends and behaviour, allowing businesses to make proactive, Knowledge- driven decisions. Businesses can make predictions about how well a product will sell or develop new advertising campaigns by using these new relationships reflected by the data mining algorithms. Data mining uses information from past data to analyse the outcome of a particular problem or situation that may arises. Data mining work to analyse data stored that data is bring analysed. That particular data may come from all parts of business from the production to the management. Managers also use data mining to compare and contrast among competitors. Data mining interrupt its data into time analysis that can be used to increase sales, promote new product, or delete that is not value added to the company.

- Offices requiring analysis or dissemination of geo-referenced statistical data.
- Public health services searching for explanations of disease clustering.
- Environment agencies assessing the impact of changing land use patterns on climate change.
- Geo-marketing companies doing customer segmentation based on spatial location.

Effects of data mining tools being delivered by Cloud are:

The customer only pays for the data mining that is need to reduce his cost for complex data mining. The customer do not have maintains of hardware infrastructure as he can apply for data mining through a browser. Using cloud computing through data mining reduces the barriers that keep small companies from benefiting of the data mining instruments. The relationship between cloud computing through the Data Mining is the cloud uses to store the data in into the server and data mining is uses to provide clients server relationship as a service and the information being collected ethical issues like privacy and individuality and violated. According to the cloud computing the security reasons are less and the data can be loss for previous reasons they use data mining for the security reason in cloud from the attackers. The attackers can use cheap and raw computing for hacking the data base in the storage of the cloud so

Figure-1

the data can be loss in server. Some mining algorithms are good enough to extract information up to the limit that violates client privacy.For example: multivariate analysis identifies. The relationship among variables and this technique can be used to determine the financial condition of an individual from his buy-sell records, clustering algorithms can be used to categorize people or entities and are suitable for finding behavioral patterns; association rule mining can be used to discover association relationships among large number of business transaction records etc. Analysis of GPS data is common nowadays and the results of such analysis can be used to create a comprehensive profile of a person covering his financial, health and social status. Thus analysis of data can reveal private information about a user and leaking this sort of information may do significant harm. As more research works are being done on mining, improved algorithms and tools are being developed. Thus, data mining is becoming more powerful and possessing more threat to cloud users. In Upcoming days, data mining based privacy attack can be a more regular weapon to be used against cloud users. In this approach we use the cloud computing through data mining in privacy reason.

IV. Conclusion

Cloud computing provides storage of data in a server by protecting data by using data mining concept. Actually we have discussed the cloud computing data mining for the advance use of security in data loss purpose. While the data we are storing in cloud is being separated in different servers for a security but the hackers using the cheap and raw cloud computing for the misuse of the software. In Cloud computing the data is being shifted from one server to another server in a peer to peer transaction As this cloud computing Data mining topic is used as a data encryption or data security from the data base.

References

1. "CloudComputingNewWine.pdf."http://www.cmlab.csie.ntu.edu.tw/~jimmychad/CN2011/Reading/CloudComputingNewWine.pdf.
2. "Top-Cloud-Computing-companies" http:www.itstrategists.com/Top-Cloud-Computing-companies.aspx.
3. Nada Elgendy, Ahmad Elragal, Big Data Analytics: A Literature Review Paper, Springer International Publishing Switzerland 2014
4. "Data mining and Analytics"- http://cseweb.http://cseweb.ucsd.edu/~elkan/255/books.html
5. "Data mining concept by Doug Alexander"-http://www.laits.utexas.edu/~anorman/BUS.FOR/course.mat/Alex/
6. "Concepts of cloud computing"-http://searchitchannel.techtarget.com/tip/Key-aspects-of-cloud-computing-services
7. "Different characteristic for the cloud computing"-deca.cuc.edu.cn/.../The-Characteristics-of-Cloud-Computing.pdf
8. "Grid Computing"-http://www.gridcafe.org/EN/what-is-the-grid.html
9. "Mainframe Computer" http://www.businessdictionary.com/definition/mainframe-computer.html
10. Praveen Kumar, Dr Vijay Singh Rathore, Efficient Capabilities Of Processing Big Data using Hadoop Map Reduce, International Journal Of Advanced Research in Computer and communication Engineering Vol. 3, Issue 6, June 2014
11. S. Rama sree, K. devi Priya, An Insight Of Big Data Analytics Using Hadoop, CSI Communications, November 2016
12. Swati Harinkhere, Nishcol Mishra, Yogendra P S Maravi, Varsha Sharma, Big Data Processing Techniques & Application: A Technical Review, CSI Communications, November 2016

RFID: A New Approach for Employee Attendance Management System

Sushma Malik*
Prashant Kumar**

Abstract

In any organization attendance of the worker play a very important role because in many organizations employee working on the hour system and payment also provide on the basis of the hour. So accurate attendance system is needed to give the wages to the employee.

Manual attendance system uses the paper sheet time cards or time sheets to confirm the attendance of the employee. The employee fills out these sheets for attendance and managers or other hired person to manage for accurateness. However, time and attendance information is issue to human error when various employees such as workers, managers, and payroll administrators all perform tasks that involve recording the numbers. Every employee punches in and punches out when they arrive in the organization.

The payroll administrator or manager record the worked hours of the employee using each employee's time card and fill the time sheet. Lots of time spent into calculating the worked hours of the workers when using the manual system. Some time the employee doing the error during the record of the attendance or during to examine the attendance or some time intentionally doing the error.

Hence, there is an essential for the payroll administrator of organization to develop the computer-based employee attendance system which will help for maintaining attendance of Employee. In this paper, we propose a RFID Based Attendance System for the organizations which provides the functionalities of the workers in organization, taking attendance, and making decision for the wages of the employee.

Key Words: RFID, attendance system, RFID Tags, RFID Readers

I. Introduction

RFID, means Radio Frequency Identification, is an automatic recognition technology used for retrieving data from or storing data on to RFID Tags without any physical contact [1]. RFID is an automated identification and data collection technology, that ensures more accurate and timely data entry.

RFID gained more attention because of its cheapness. RFID is a combination of Radio frequency and microchip technology to construct a smart system which is used to identify, and monitoring the record [9].

Sushma Malik*
Assistant Professor, IITM, Janak Puri, New Delhi
Prashant Kumar*
Assistant Professor, Banarsidas Chandi Wala Institute of Professional Studies

The main components of an RFID system are:

- RFID tags
- RFID Reader
- Middleware
- Backend Database

The tag consists combination of zero and ones which is called the bits, and the transmission and receiving of data has a communication channel. The tag can be an electronic device with its own power supply is called an active tag or a very low power integrated circuit is called passive tag which takes energy from the scanner to transmit its content. The reader is the most fundamental part of the RFID system.

It reads data from the tag and transmit it to the Middleware for additional processing [12]. The reader

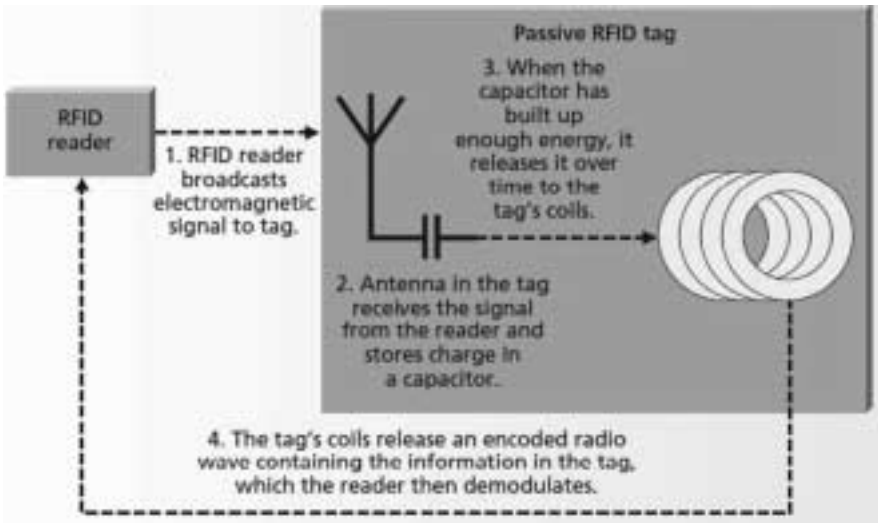


Figure 1. Simplified view of data transfer in low frequency passive RFID tags [13]

attempt to cross-examine the tags at varying frequencies. The reader communicates to each other by transmitting a beam of impulses, which encapsulate commands to the tag and listens for the tag's response [10].

The reader is connected to the workstation for data processing with wire or over a wireless connection. The Middleware is an interface that encompasses all those components that are responsible for the transmission of data from the reader to the backend database management systems [8].

The Middleware can include hardware components like cables and connectivity ports and software components like filters that monitor network performance of the system [2, 9]. The backend database primarily deals with the storage of relevant information recorded by the reader and communicated by the middleware [12].The Backend database stores individual tag identifiers to uniquely identify the roles of each tag.

The database stores record entries pertaining to individual tags and its role in the system application.

The RFID system is mutually dependent on its main components to achieve maximum efficiency and optimum performance of the application. Due to its high degree of flexibility, the system can be easily adopted for the attendance system in the organization.

II. Review of Related Works

A number of related works exist in literature, application of RFID Technology to different areas and specifically to the area of academic attendance monitoring problem. Mohamed A.B, Abdel-Hamid A and Mohammed K.Y authors designed and implemented a model of a secured and portable embedded reader system to read the biometric data from the electronic passport.

The authors attempted to solve problems of reliability, security and privacy in E-passports by authenticating holder online using Global System of Mobile Communications (GSM) network. The GSM network is the main interface between identification centre and the e passport reader. The communication data is protected between server and e-passport reader by using AES to encrypt data for protection while transferring through GSM network [14].

C. E. Geoffrey said, the automatic access control system evolves to avoid illegitimate entry of people into a building and preventing unlawful people from gaining access to certain organization resources. The system also stores the login and logout information of the user [16]. K. D. Mahajan, P. Pandey and B. K.

Pandher, applied RFID Technology in Libraries and according to them RFID based library management

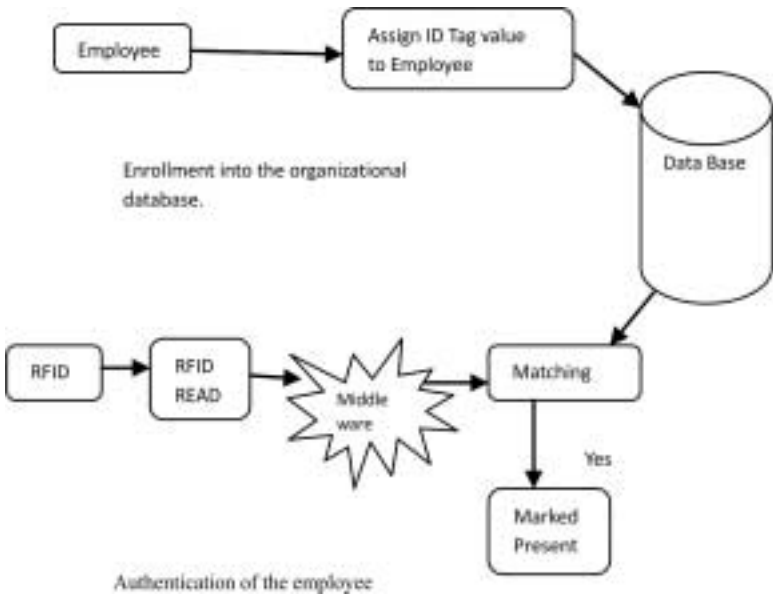


Figure 2: General architecture of a proposed system

saves the library staff's time and energy by automating their task. borrow and return of books are automated using the check-in, check-out system which is RFID based. The limitation of that system is that it is costly to implement [18].

III. Proposed System

In this paper we mainly focus on the new approach to take the attendance system in the organization. It includes the two phases: Enrollment into the organizational database. Authentication of the employee for the attendance.

Enrollment phase is basically used for the registration of the new Employee when hired in the Organization, in the Database. It stores the ID tag associate with the Employee in the database.

During Authentication phase, ID Tag matching is performed to check the validation of the Worker in the Organization.

IV. Conclusion

RFID used retrieve data from the stored database and compares that data to RFID Tags without any physical contact.

RFID is data collection and automated identification technology which is used for accurate and timely data entry.

This paper introduces the automatic attendance system trough RFID. This system allows employee to simply fill their attendance just by swiping or moving their ID cards above the RFID reader which are located at the entry point of the organization for work.

We hope that this system can shift the old method of attendance system to new computerized attendance system and provide a new, accurate, and less time consuming way of taking attendance of the employee.

References

1. L. Sandip, "RFID Sourcebook", IBM Press, USA, (2005) ISBN: 0 -13-185137-3.
2. E. Zeisel & R. Sabella, "RFID+", Exam Cram, (2006), ISBN: 0 -7897-3504-0. [3] US.Department of Homeland Security, "Additional Guidance and Security Controls are needed over Systems using RFID and DHS", (2006), OIG-06-53.

3. US. Department of Homeland Security, "Enhanced Security Controls needed for US-Visit's System using RFID Technology", Department of Homeland Security (Office of Inspector General), (2006), OIG-06-39.
4. US. Government Accountability Office, "Information Security: RFID Technology in the Federal Government", (2005), Report to Congressional Requesters, GAO-05-551.
5. K. Ahsan, H. Shah, P. Kingston, "Role of Enterprise Architecture in healthcare IT", Proceeding ITNG2009, (2009), IEEE.
6. Intermec, "ABCs of RFID: Understanding and using radio frequency identification", White Paper, (2009).
7. Juels, A., Weis, S.A. (2005). Authenticating Pervasive Devices with Human Protocols. Advances in Cryptology – Crypto '05. Lecture Notes in Computer Science. Volume 3621. pp 293-308.
8. Arulogun O. T., Olatunbosun, A., Fakolujo O. A., and Olaniyi, O. M.(2013)," RFID-Based Students Attendance Management System", International Journal of Scientific & Engineering Research Volume 4, Issue 2, February-2013 1ISSN 2229-5518
9. RFID Journal. (2003). Gillette Confirms RFID Purchase. RFID Journal. Available at: http://www.rfidjournal.com/article/articleview/258/1/1/.
10. Krane, J. (2003). Benetton clothing to carry tiny tracking transmitters. Associated Press.
11. Albrecht, K., and McIntyre, L. (2005). Spychips: How Major Corporations and Government Plan to Track Your Every Move with RFID. Nelson Current Publishing
12. Shashank Shukla, Shailee Shah, Pooja Save (2013), RFID Based Attendance Management System, International Journal of Electrical and Computer Engineering (IJECE) Vol. 3, No. 6, December 2013, pp. 784~790 ISSN: 2088-8708
13. Mohamed A.B, Abdel-Hamid A and Mohammed K.Y.,(2009), "Implementation of an Improved secure system detection for E passport by using EPC RFID tags", World Academy of Science, Engineering and Technology Journal, Volume 6,pp1-5.
14. RFID Sens Net Lab (2005), A white paper on Automatic Attendance System. Texas A & M University, Texas, USA.
15. C. E. Geoffrey, "Automatic Access Control System using Student's Identification Card based on RFID Technology", Unpublished Thesis Faculty of Electrical Engineering, University of Teknologi Malaysia, (2012).
16. A. Parvathy, V. R. Raj and M. Reddy, "RFID based exam hall maintenance system", IJCA Special Issue on Artificial Intelligence Techniques-Novel Approaches & Practical Applications (AIT). (2011).
17. K. D. Mahajan, P. Pandey and B. K. Pandher, "Application of RFID Technology in Libraries and Role of Librarian", (2010), pp. 109-118
18. Kommu Ayyanna, "study of some fingerprint verification algorithms", National Institute of Technology, Rourkela

Virtualization: A Boon for Indian Rural Education System

Kanika Jethwani*
Kalpana Jha**

Abstract

In villages not much attention is given to education probably this is due to lack of awareness. India has diverse population and limited resources where providing education to the Children of the nation is the major concern. In order to increase Villager's interest in education government is providing free books, mid day meal scholarships and other benefits.This paper throws light on Education in rural India; the challenges faced and discuss the concept of virtualization in education which can help rural India to get better education.

Key Words: Virtualization, Virtual Classroom, Virtual Machine.

I. Introduction

India's real assets are the children who are in schools; they are going to make the country proud. But the problem is that people in rural India can't afford donations taken by public schools and unfortunately the government schools in rural India have failed in providing quality based education. Education components needs to be standardized with quality based parameters. There are a number of prominent factors that are a concern [1].

- One of the issues of concern is that the present rural schools are degraded. Despite of the fact that 60% of the children living in rural areas are enrolled in schools and 70% of the public schools are there in rural areas, still there is lack of quality education as well as the quality of teachers is not upto the mark.
- Second issue is the lack of basic infrastructure. The schools are not equipped with basic facilities such as classroom, electricity etc.
- Third issue is the quality of study material which is obsolete and needs to be upgraded. The course content is theoretical and along with this the vocational aspect still remains uncovered [3]. The students do not get practical exposure to the concepts and hence are unable to understand technologies properly.
- Fourth issue is the lack of awareness among rural India about the importance of education. Since people in rural India are unaware about the need of education and its benefits, they remain uninterested in gaining knowledge by going to school.

The world has become technical, education is no different. The need of hour is to get anywhere, anytime education. Virtual education is the key which allows students to get classroom environment virtually. There are several steps that the government of India is taking for the up gradation of education in rural India and the use of technology will largely help in imparting knowledge in most efficient manner. Hence virtualization can largely bring revolution in rural Indian education system [2].

II. Virtualization

Virtualization means creating virtual instance of the hardware which behaves like system with an operating

Kanika Jethwani*
Asst. Professor, IT
IITM, Janakpuri

Kalpana Jha**
Asst. Professor, IT
IITM, Janakpuri

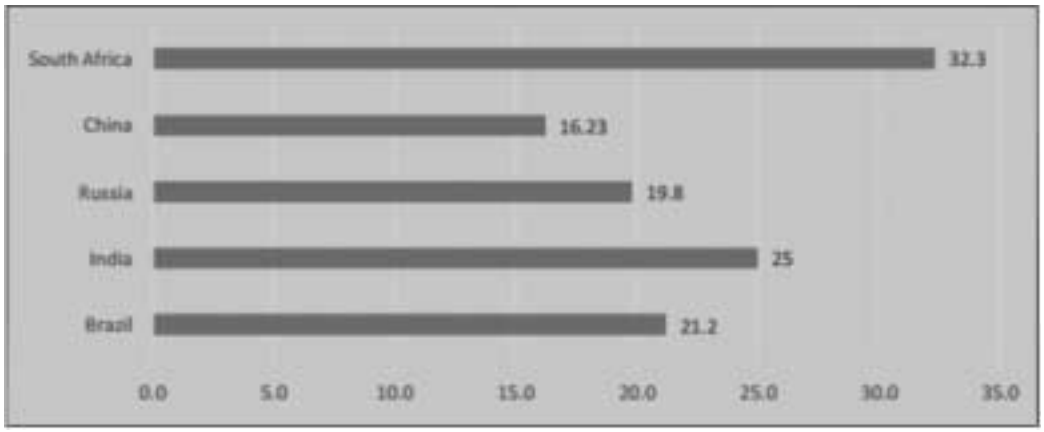


Figure 1: Pupil Teacher ratio in primary Education during 2014 (mhrd.gov.in)

system. It basically describes separating resources and requests for any service from the delivery of that service from the delivery of that service physically. For example, with the virtual memory a system can acquire more memory than the actual installed memory through swapping data to disc storage.

Virtualization has its applications in large spectrum of fields such as entertainment, architectural design, education, media and medical treatment. It provides the blend of various technologies wherein there is a layer of abstraction between processor, storage and hardware.

The computing era has changed after the evolution of virtualization. We look at the comparative view of before and after virtualization [4].

Before Virtualization

- There was single operating system on a machine.
- Both hardware and software were tightly coupled.
- Running of more than one application on a same machine created conflict.
- Resources were not fully utilized.
- Infrastructure was inflexible and costly.

After Virtualization

- The operating system and applications have hardware independence.
- Virtual machines can be put into any machine.

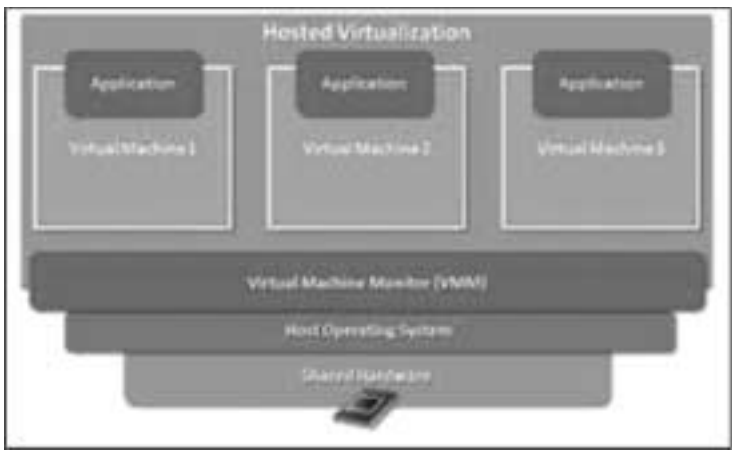


Figure 2: Virtualization Architecture

- Operating system and applications can be encapsulated into a single unit into virtual machines.

III. Architecture

In virtualization, there is a host system wherein the virtualization takes place and there can be one or more virtual machines. If the host machine is running Microsoft OS and you want your system to behave like it is running Linux OS then you can run that OS on virtual machine and your system looks like a system with Linux OS.The software which creates a virtual machine (VM) on whose system hardware is called Virtual Machine Monitor.

A. Meaning of Virtual Machine

A Virtual Machine is an emulation of computer system similar to physical machine that runs Os and applications. A virtual machine comprises of a set of configuration and specification files which are backed by physical resources of the host machine. A Virtual Machine uses the hardware more efficiently which helps in maintaining costs and reducing power and cooling demand.

Each Virtual Machine has a virtual device which provides all functionalities that are provided by the physical machines with various other benefits such as security, portability and manageability.

IV. Virtual Classroom Education

A. Meaning of Virtual Education

A Virtual Classroom of a school or institute is based on the usage of modern technologies for teaching learning pedagogy that helps in increasing the knowledge and its discernment. Here the teacher and student may be separated by distance, time or both and the teacher can easily provide the content of course through multimedia resources, videoconferencing and the Internet etc. The virtual education is boom for the environment as it reduces the use of paper and helps in saving the environment by keeping the data safe in the computer.

B. Virtual classroom

Virtual classroom is a teaching learning environment where students and teacher interact, view and discuss presentations or videos and engage with resources and groups.

There is a huge demand of Indian teacher not only in India but throughout the world. Here the students of India travel abroad to pursue higher education with the help of new technology and knowledge sharing platform both in India and abroad.

The Government of India has taken many steps to reform the teaching pedagogy in India by inculcating the concept of virtualization within the teaching learning patterns. “Coaching practices is growing highly commercial in India, preparation for entire examination for IIT, IIM and other institute of repute attract 60 lacks application, who spend an average of nearly 12000 each on coaching amounting rupees 7200 crore per year. [1]. Many coaching institutes serves its students through online and correspondence courses, that eventually emphasis the need of virtual education in India. A virtual classroom enhances the learning with its interactive features like audio-video, whiteboards, chatting, content sharing etc.

A virtual classroom can broadly be categorized as

- Unsupervised Virtual Classroom- It does not require the need of an instructor as it consists of ready-made study materials that students can make use of as it has self-paced tutorial course where the exams are conducted automatically after completion of every activity. Here students can make use of a PowerPoint presentation or a video tutorial.
- Supervised Virtual Classroom- It involves a traditional classroom environment where at least one instructor is present and the teaching is done in real time and date, with the students are present virtually through any video conferencing application. Here, interaction and participation among teacher and students is active.

Benefits of a virtual classroom

- Students can get benefit of the study material available worldwide.
- Interactive learning seeks student’s attention.

Limitations of a virtual classroom

- Supervised classes require a fixed schedule to be followed by the students which may be a constraint for some of the students.
- Up gradation of hardware and software as well as the Internet speeds may be a limitation.

V. Conclusion

India is a country that has huge amount of skill and brain that is a worldwide demand. The country has huge amount of knowledge among the people. Since past Takshila and Nalanda were the oldest and most developed universities throughout the world. Today, the government is also taking a lot of efforts to enhance the quality of education. For this, the government keeps on announcing and imposing many rules and regulations in the education sector, be it primary, secondary or higher education. For the enhancement and up gradation of education, many steps have been taken by making use of ICT and virtualization. So far it has also been successful to virtualize the education system in many of the universities and institutes like IIT, IIM, JNU, AIIMs etc. However, there are many more challenges and hindrances to be overcome to achieve virtual education throughout the nation. Virtualization can bring boon to Indian rural education system by its effective applications.

References

1. Deshwal, A. and Dobhal, A., "Virtualization: A Tool to a Sustainable Education System", IJARCSSE, 3, March, 2013, pp. 539-544
2. Dinesha, H., A., and Agarwal, V., K., "Advanced Techniologies and tools for Indian Rural school Education System", IJCA, 10, December, 2011, pp. 54-59.
3. Dwivedi, A. and Dheeran, A., "Virtualization of Education in Modern Era".
4. VMWare Whitepaper, "Virtualization Overview"

Blood Mate – An Android Application to Illustrate the Perspective of Community on Blood Donation

Kavita Pabreja*
Akanksha Bhasin**

Abstract

Blood is the supreme element in any individual's life support system. It is the elementary unit in the existence of mankind without which the living being's survival in this world is impossible. Additionally, the process of blood formation takes place only by natural means i.e. it can neither be synthesized nor be produced artificially in any laboratory. Hence, the only alternative left for the fulfilment of various blood needs and requirements is through blood donation. Also, people are less aware about the characteristic properties of their blood. The destitution of this application is felt in an attempt to generate awareness among people about the same and discuss the various traits that are specific to each blood group. As the name says, "Blood Mate", it certainly explains to the user every possible detail about his blood group largely in simple and basic terms that can be well understood by any individual. The major features of this app embraces details about each blood group based on diet, personality and career choices that can be taken by an individual, supplying various tips for blood donation and predicting the blood group of an unborn child as well The present paper is based on an attempt to investigate the various features of the android application named "Blood Mate" developed by authors.

Key Words: Android Studio, Global Positioning System, Software Development Kit, System Analysis

I. Introduction

In India, October 1st is observed as the 'National Blood Donation Day.' The red stuff that oozes out of a person's body after a paper cut or logically speaking, a fluid that circulates constantly in our body, providing us with the basic nutrition, oxygen and waste removal, is called as Blood. As explained by the author in [1], it is not just a simple fluid, but a group of cells suspended in it along with proteins that not only support our life system but also enables us to fight diseases and infections. Blood can neither be manufactured nor there do any substitute available for it. Hence, Blood Donation is the only source through which blood can be received. As stated by the author [2], blood group system was discovered in 1901 by Karl Landsteiner. So far 19 major groups have been identified of which "ABO" and "Rhesus" groups are of major importance. The genetics of blood groups is proved by the fact that specific diseases are common in particular blood group; for example: duodenal ulcers in 'O' blood group, gastric cancer in 'A' blood group.

Kavita Pabreja*
Ph.D., Associate Professor, Head of Department of Computer Science Maharaja Surajmal Institute, New Delhi-110058
Akanksha Bhasin**
Student, Bachelor of Computer Applications Maharaja Surajmal Institute, New Delhi-110058

A. System Analysis:

Analysis can be defined as breaking up of any whole so as to find out their nature, function etc. It defines design as to make preliminary sketches of; to sketch a pattern or outline for plan, to plan and carry out especially by artistic arrangement or in a skilled manner. System analysis and design can be characterized as a set of techniques and processes, a community of interests, a culture and an intellectual orientation. The various tasks in the system analysis include the following.

- Understanding application.
- Planning.
- Scheduling.
- Developing candidate solution.
- Performing cost benefit analysis.
- Recommending alternative solutions.
- Supervising, installing and maintaining the system.

This system provides the information about the diet specific to each blood group along with the ability to distinguish between the various personality traits amongst the blood groups, predicting the possible career choices as well as finding the blood group of an unborn child. The system also aims at the inclusion of various data mining results so as to generate results that can further be added to the success of the app. Along with this; general blood donation tips have also been provided. The system is interactive, and has an easy to use interface.

B. Existing Android Apps:

The existing and available applications on the Google Play Store deal with only a single or at most two of the factors. This leads to the problem for the user as he/she has to download a variety of applications which can lead to a lot of memory wastage over and above filling the mobile phones with a lot of useless material. But apparently, there is no common platform for users to access all facilities together. This creates a need for the android users to download multiple apps and surf each one separately for each of their functionality. In conjunction, people have to consult various dieticians and nutritionists so as to maintain a healthy diet as well as sustaining their well-beings which steers to wastage of time and money. This is the chief shortcoming of the existing systems.

C. Proposed System:

To overcome the weaknesses and problems that lie in the existing android application, this project has been evolved. It aims to reduce the time and money wastage by providing an integrated platform for all the modules. It also targets to provide precautionary measures to be taken, by advising the blood donation tips through a simple application. The system or app provides an easy to use interface. The various other advantages of the proposed system are:

- Trouble free to use
- Relatively faster than traditional methods
- Highly reliable
- Easy and interactive GUI
- Easy to operate and maintain
- Use of latest technological means

II. Literature Review

The applications that are currently present on Google Play Store provide information about a single attribute associated with blood. Instead of downloading multiple applications and wasting space, "Blood Mate" aims to deliver a single integrated platform for all facets related to blood. Following is the list of reviewed applications-

A. Blood Groups and You

This application [3] supplies information about only two aspects *viz.* the blood donation tips along with prediction of the blood group of the unborn child based on the input from the parents i.e. taking their blood groups' as an input.

B. Diet-

This Android application [4] imparts details about only one attribute i.e. about the diet that is specific to each blood group. It does not support a good user interface as it consists of a slide show comprising of 11 screens that describe about the diet pertaining to each blood group.

C. Blood Type

It also incorporates [5] information about the diet for each group besides letting the user know from which group he can receive blood and to which group he can donate blood i.e. the possible donor-receptor groups.

As evident, this creates a need for the android users to download multiple applications and surf each one discretely and exclusively for each of their functionality that shams as the major drawback of the prevailing system. Hence, Blood Mate caters to this pre-requisite

by offering a unified platform to overcome the abstruseness and deficiencies of the above mentioned applications accompanied with additional features in details.

III. Methodology Used

The software(s) needed to develop this application are as follows:

A. Android Studio -

Android Studio provides the fastest tools for building apps on every type of Android device. World-class code editing, debugging, performance tooling, a flexible build system, and an instant build/deploy system all allow you to focus on building unique and high quality apps. Android software development is the process by which new applications are created for the Android operating system. Applications are usually developed in Java programming language using the Android software development kit (SDK), but other development environments are also available. The Android SDK includes a comprehensive set of development tools. These include a debugger, libraries, a handset emulator based on Quick Emulator, documentation, sample code, and tutorials. Currently supported development platforms include computers running Linux (any modern desktop Linux distribution), Mac OS X 10.5.8 or later, and Windows XP or later. Following are the features of Android Studio:

- Instant Run
- Intelligent Code Editor
- Fast & Feature Rich Emulator
- Robust & Flexible Build System
- Designed for Teams
- Optimized for all Android Devices

B. Java Development Kit

The Java Development Kit (JDK) is an implementation of either one of the Java Standard Edition, Java Enterprise Edition or Java Micro Edition platforms released by Oracle Corporation in the form of a binary product aimed at Java developers on Solaris, Linux, Mac OS X or Windows. The JDK includes a private Java Virtual Machine and a few other resources to finish the development of a Java Application. Since the introduction of the Java platform, it has been by far the most widely used Software Development Kit.

C. Adobe Photoshop -

Photoshop is a graphics-editing program that is used to create and manipulate images. The program's versatile nature makes it useful for a huge range of imaging tasks. This software has been used in this project to creatively design a simple but attractive and efficient user interface design for the users. All tools present on the app screen have been designed using Adobe Photoshop.

IV. Project Insights

'Blood Mate' android application deals with postulating a cohesive and amalgamated platform for various applications that are present on Google Play Store and incorporating the features of all those applications into a single component so as to ease the glitch of numerous downloads of diverse applications thereby condensing the wastage of memory space and time. Likewise it stipulates the details in a friendly and simple language so that any individual can understand it in an utmost convenient and competent manner. The application can currently be divided into five modules. The authors have created a dashboard which includes the assimilation of the following modules:

A. Blood Diet

The module named as 'Blood Diet' which provides information on the diet to be followed by a particular blood group.

B. Donor- Receptor Groups

The module named as 'Donor-Receptor Groups' which provides the information on the possible recipient and donor blood groups i.e. to which blood group the user can donate the blood and from which blood group he/she can receive blood.

C. Explore yourself

A module named as 'Explore Your Self' that discusses about the various personality traits and career choices that are usually similar in characteristic blood groups.

D. Your Baby's Blood Group

The module named as 'Your Baby's Blood Group' which helps in predicting the blood group of the unborn child by taking the father's and mother's blood groups as an input.

E. About Donating Blood

The module named as 'About Donating Blood' which provides tips for effective blood donation as well as precautions that can be taken.

V. Results and Discussions

The following screenshots of the application demonstrates its various modules:

- The "Diet" module consists of a set of radio-buttons which on clicking displays the list of food items that a particular blood group should follow, as shown in Fig.1.
- The "Baby Blood Group" module entails 2 radio-groups which on selecting a blood group from each group, displays the probable blood group of the unborn child in a text view, as shown in Fig.2. It uses a set of nested if-else constructs where the input is taken from the 2 radio groups; in which each radio group contains the blood group of the child's parents i.e. the mother and father. The input obtained is then used for finding the blood group of the child and is then displayed in a text view.
- Fig. 1 demonstrating the "Diet" activity which includes a set of radio-buttons based on the blood group of a person

Fig. 1 "Diet"

Fig. 2. "Baby Blood Group"

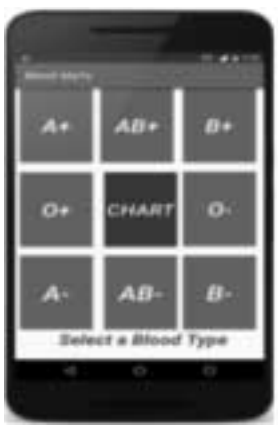


Fig.3 "Donor-Receptor Groups"

Fig.4 "About Donating Blood"

The "Donor-Receptor Groups" Module shown in Fig.3 consists of a table of buttons that when clicked, produces the output as the blood groups to which the user can donate and from whom they can receive.

The "About Donating Blood" module that comprises of a list-view which provides various blood donation tips in the form of a list, as shown in Fig.4

VI. Conclusions and Future Scope

To conclude, "Blood Mate" has a very wide scope in terms of generating awareness among people for understanding the most important aspect of their health – 'Blood.' In consort with its various modules, it also delivers an easy implementation environment and generates responses flexibly.

In the future, the user would be able to use Global Positioning Systems (GPS) services so as to aid him/her in locating the nearest clinics and blood banks and also provide them lab facilities at their homes itself. This facility would be chargeable and would offer rewards and perks to donors.

Also, the app aims to further integrate the finger-print sensing technology with which the blood group of an individual can be predicted based on the fingerprint's pattern.

It is also proposed to extend this study by collecting data from students of various under-graduate courses so as to get a real picture of blood donations and associated beliefs. Data, hence collected, would be mined for extracting hidden facts related to blood donation.

Fig. 4 illustrating the "About Donating Blood" module that provides a list of blood donation tips

References

1. Adobe Photoshop review from: http://kivaindia.com/Photoshop-Institute-In-Ahmedabad-Photoshop-is-a-Graphics-Editing-version-program-that-is-used-to-create-and-manipulate-images-The-program-s-versatile-/b961#
2. Android Application present on Google Play Store: https://play.google.com/store/apps/details?id=it.betterdays.dietagrupposanguignoandroid&hl=en
3. Android Application present on Google Play Store: https://play.google.com/store/apps/details?id=blood.ciencia.technologies&hl=en
4. Android Application present on Google Play Store: https://play.google.com/store/apps/details?id=malix.diet&hl=en
5. Android Studio review from: https://developer.android.com/studio/index.html
6. F. Haniff, "Blood and its importance."www.kaieteurnewsonline.com, March 2009.
7. P. Narkhede, "An empirical study on Blood Types and Personality." IJSSBT, 2 June 2015.
8. Java Development Kit review from: https://en.wikipedia.org/wiki/Java_Development_Kit

Software Testing- A Brief Study

Shailja Mehta*
Himanshi Bhambri**

Abstract

Software development has constantly been a task and much over becomes when the software needs to be incorrigible. Software testing is an indispensable task in the ocean of production of software's where it plays its basic role by accomplishing. Through this paper, we are trying to present a brief study on software testing by focusing on its objectives and levels. Also, here we bring in light, the various challenges and types of software testing. Finally the paper is concluded with a contrast between the types of testing.

Key Words: Verification, validation, stubs, drivers, white-box, black-box, grey-box.

I. Introduction

Software testing is an activity done with intent of determining errors within a given snippet. In this process, we begin executing a program or application with the aim of finding the maximum possible errors, so that the delivered product can gain customer trust and confidence. It can also be quantified as the route of validating and verifying that a software program or application encounter and meet the business and technical requirements that guided its design and development. Software testing is not a "silver bullet" that can guarantee the production of high quality software systems as it clearly concluded that absence of errors do not signify that software is bug-free which basically concludes, "A correctness proof demonstrates that a software system which exactly meets its specification will always operate in a given manner, software testing that is not fully exhaustive can only suggest the presence of flaws and cannot prove their absence [1]." Software testing is basically performed by software testers at different levels. Software testers need to understand the software well in order to determine the variation. Since it is much unspoken that to claim error in somebody's work is not easy, so such parameters make software testing a difficult task. Software testing provides a means to reduce errors, cut maintenance and overall software costs. To alleviate testing procedure, many new techniques, tools, methodologies have been developed in the recent times which not only enhances the software quality but also improves the throughput. The two basic terms related to software testing are Validation and Verification which lay the basic framework for the above said activity. Testing in itself is not a single world but multiplicity of tasks starting from planning, designing test cases, implementing test cases, preparing reports and much more. Software testing is a broad term encompassing a wide spectrum of different activities, from the testing of a small piece of code by the developer (unit testing), to the customer validation of a large information system (acceptance testing), to the monitoring at run-time of a network-centric service-oriented application [3].

Shailja Mehta
IINTM, Janakpuri
E-15A, Uttam Nagar, New Delhi

Himanshi Bhambri
Assistant Professor-IT
D-27, 28, Janakpuri, New Delhi

II. Objectives of testing

Software Testing is a process of finding errors while executing a program so that we get a zero defect software [2]. Despite of testing being difficult, it is performed with multiple aims as listed:

i. Software Testing establishes that the software is working acceptably as per the requirement specification. This means that we are confirmed that we have built the right software.

ii. It consists of successions designed and implemented to prove that the program is error free. In many phases through which we pass during testing process, new test cases are developed and performed to discover maximum possible errors.

iii. It is the job of testing to certify that the software does its job correctly and can be used in production. Whenever software passes through the testing phase, a testing report is prepared at the end containing the errors discovered and solved. Once the final report is given it automatically invests the customer's trust to use the software commercially.

iv. The tenacity of testing is to show that the software works.

III. Levels of testing

Testing is not a flat process to be performed. Software testing is more than just error detection [4]. In every phase of software development, testing is performed in its own way. In a broader sense, it is divided into various levels as stated below:

A. Unit Testing

This sort of testing is performed majorly by developers who are the creators of their own code. It is done before the setup is handed over to the testing team to formally implement the test cases. Unit testing is performed by the respective developers on the individual units of source code developed to check whether the unit is performing the intended task in the expected manner. The developers must use test data that is different from the test data of the quality assurance team who is ultimately going to perform the testing. The goal of unit testing is to test each part of the program in an isolated manner and show that individual parts are correct in terms of requirements and functionality. Also, we can say since here the developer is dealing with his own piece of code so, he can perform a well defined functional and structural testing.

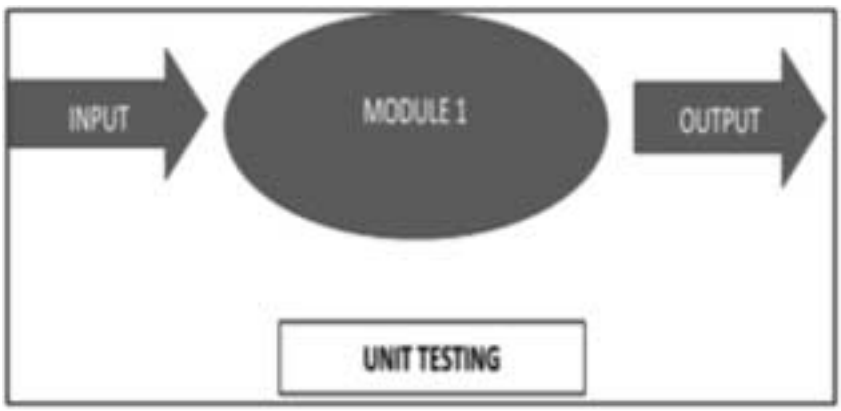


Fig. 1. Diagram showing the concept of Unit Testing.

B. Integration Testing:

Integration testing is acknowledged as uniting several independent parts of code and then performing the process of software testing. In other words, it is the testing of several combined parts of an application to determine if they execute correctly and communicate effectively by giving appropriate results. Integration testing can unite codes in any manner but is performed in two ways: Bottom-up integration testing and Top-down integration testing. Sometimes the concept of integration testing includes the terms like drivers and stubs. Drivers and stubs are nothing but the substitutes to fill the blank spaces during integration testing for the currently missing or parts under development. This can be best explained as follows:

- **Bottom-up integration**

In this type of testing, we generally begin testing from the lowermost codes and then proceed upwards. When we begin from the base level then we need some codes to call them in order to check their functionality. In

such a situation, if the codes on the upper layer are available or designed till that point of time then they make a call to the succeeding ones otherwise dummy superiors are used known as drivers to make a call to the juniors. The application is tested thoroughly to verify that it meets the functional and technical specification. The application is tested in an environment that is very local to the production environment where the application will be deployed. So here, we basically aim towards testing from the lowermost layer towards the topmost.

- **Top-down integration**

In this type of testing, we generally begin testing from the uppermost codes and then proceed downwards. When we begin from the top level then we need some codes to be called by them in order to check their functionality. In such a situation, if the codes on the lower layer are available or designed till that point of time then they can make a call to the succeeding ones otherwise dummy juniors are used known as stubs . The application is tested thoroughly to verify that it meets the functional and technical specification. The application is tested in an environment that is very local to the production environment where the application will be deployed. So here, we basically aim towards testing from beginning towards the lowermost layer. Thus, in this testing, the highest-level modules are tested first and progressively, lower-level modules are tested thereafter.

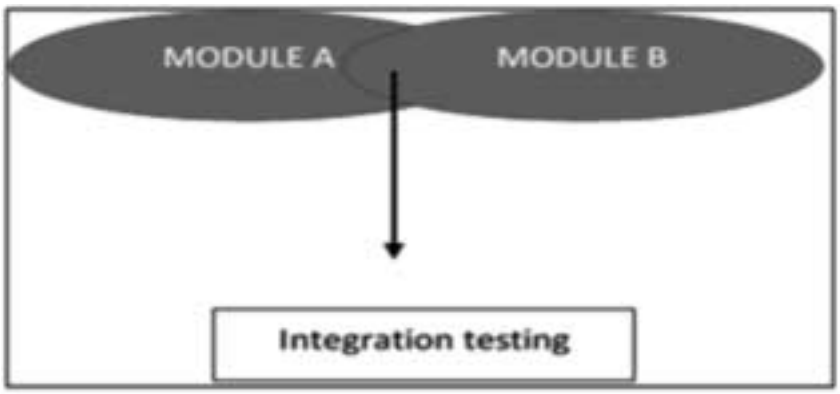


Fig. 2. Diagram showing the concept of Integration Testing.

C. System testing

System testing tests the system as a whole. Once all the components are integrated, the application as a whole is tested rigorously to see that it meets the common types like stress testing, load testing, acceptance testing but on a broader version, we specified Quality Standards. This type of testing is performed by a specialized testing team. During such type of testing, proper test reports are developed stating the lines of errors or variations which is further discussed in scheduled meetings with the developers.

IV. Approaches of testing

Testing has always been an implicit activity since the software development started. Its importance and value cannot be disregarded. Earlier and recent times have shown a lot of revolution in the field of testing by enhancing the way it was introduced and performed. Testing approaches can be classified as manual and automated which are discussed below in detail.

A. Manual Testing

Manual testing embraces testing software manually, i.e., without using any automated tool or any script. In this type, the tester behaves like an end-user and tests the software to identify any unexpected deeds or bug. It is basically a type of informal approach where the tester does not follow any rigorous testing procedure, but rather explores the user interface of the application using as many of its features as possible or using information gained in prior tests to intuitively

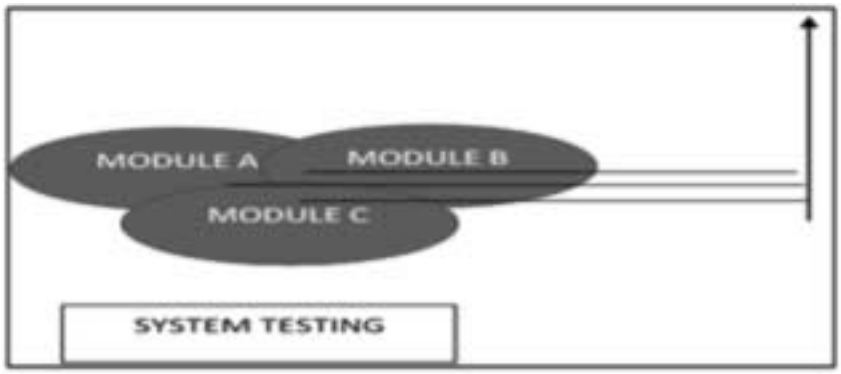


Fig. 3. Diagram showing the concept of System Testing.

derive additional tests. The success of exploratory manual testing relies heavily on the domain expertise of the tester, because a lack of knowledge will lead to incompleteness in testing.

B. Automation Testing

Automation testing, which is also discerned as Test Automation, is when the tester writes scripts and uses software to test the product. This process comprises automation of a manual process. Automation Testing is basically an automatic or software based form of testing in which software tools are used to create, execute and run the test cases. It not only provides a platform for testing but also enables for automatic reporting of errors, comparison of test cases and other activities pertaining to testing. It provides us to run the test scenarios that were performed manually, quickly, and repeatedly. This sort of testing is also used to test the application from load, performance, and stress point of view. It build ups the test coverage, improves accuracy, and saves time and money in comparison to manual testing.

V. Types of Testing

Like every phase has its own culture similarly, the beauty of testing phase is also described by the varied types. Although we have heard about many about many common types like stress testing, load testing, acceptance testing but on a broader version, we basically classify them as white-box, black-box and grey-box testing which are better elucidated below.

A. Black-box

Black box testing is a testing which forbids the tester from knows about of the internal functionality. The tester only needs to perform the functional testing as just to check the performance as expected. In Black box testing the testers have no knowledge of how the system or

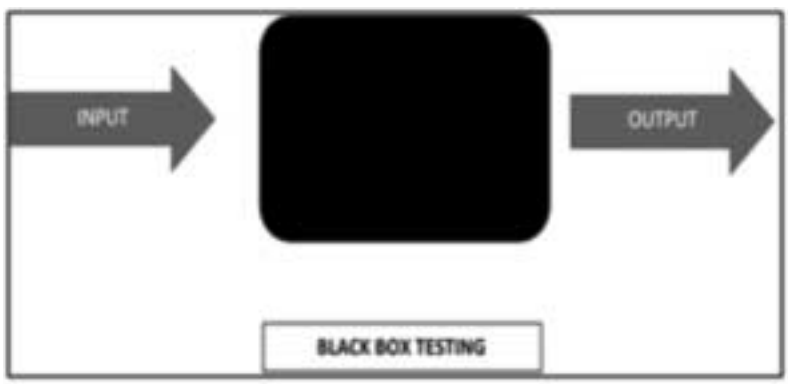


Fig. 4. Diagram showing the concept of Black box testing .

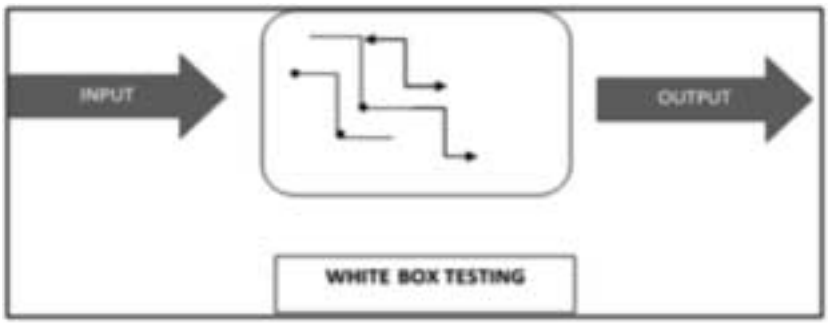


Fig. 5. Diagram showing the concept of White box Testing.

component is structured inside the box. Here, tester and programmers are independent of each other.

B. White-box

In white-box testing, the technique used requires a thorough understanding of the code insights. This means that unlike black box testing where internal knowledge of code was not necessary to the tester, here, it is the foremost necessity for the tester. Unlike black box testing, white box testing uses specific knowledge of code to examine outputs.

C. Gray box

Gray box testing is a software testing methodology which is a combination of two above mentioned strategies. It refers to a testing system by knowing limited information about the internals of the system. A little knowledge of the system is expected in Gray box testing. Here, Testing is based on the perspective of the user, rather than the designer.

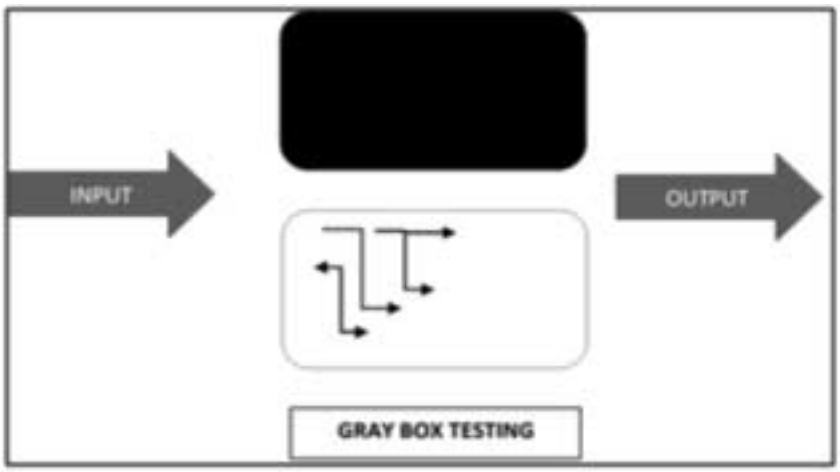


Fig. 6. Diagram showing the concept of Gray box testing.

VI. Challenges of Software Testing

- Complete testing is not possible
- The testing always should be under time constraint.
- The requirements keep on changing.
- There can be some lack of resources, tools and training
- Requires skilled set of human resource.
- Missed defects
- Out of box thinking
- Manual testing demand skills because without skills the tester might falter
- Challenges due to clients and management

VII. Comparison between Testing Types

Table 1. Table showing the concept of black box, white box, gray box testing.

S.No.	White-Box	Black-Box	Gray-Box
1.	Testers have full knowledge of internal structure and inside working.	Internal work knowledge is not required.	Somewhat knowledge of internal working are required and should be known.
2.	Done by testers and developers.	Done basically by end users and also by testers.	Basically done by testers with involvement of developers.
3.	Most time consuming and exhaustive	Least time consuming and exhaustive.	Partly time consuming and exhaustive
4.	Also known as code based testing or structural testing.	Also known as closed testing or behavioral testing.	Also known as translucent testing.

VIII. Conclusion

The study above thus states that testing is an inevitable phase for any software and not only but one of the reasons behind the success of a software or an application. Testing has also progressed so far with the evolution of new techniques and tools and also its advancements towards automated ones. Thus software testing suffices its worth towards testers, developers and end-users. Therefore we can say, "It is the process where the system requirements and system components are exercised and evaluated manually or by using automation tools to find out whether the system is satisfying the specified requirements and the differences between expected and actual results are determined [5].

References

1. A. Bansal (2014), "A Comparative Study of Software Testing Techniques," International Journal of Computer Science and Mobile Computing, 3(6), 579-584.
2. R. K. Chauhan, I.Singh, "Latest Research and Development on Software Testing Techniques and Tools," International Journal of Current Engineering and Technology, 4(4),2014.
3. A. Bertolino, "Software testing research: Achievements, challenges, dreams," In 2007 Future of Software Engineering (pp. 85-103). IEEE Computer Society, 2007.
4. S. H.Trivedi, " Software testing techniques," International Journal of Advanced Research in computer science and software Engineering, 2(10), 2012.
5. I. Hooda, R.S. Chhillar, "Software Test Process, Testing Types and Techniques," International Journal of Computer Applications, 111(13), 2015.

A Comparative Study of Social Network Analysis Tools

Hanish Chauhan*
Surbhi Kakar**

Abstract

A social network is a network which connects people to communicate with each other. It can also be defined as a structure which is used to do interaction between individuals/groups. The study of this structure is known as social network analysis. Social Network analysis is used to identify local and global patterns and examine network dynamics. This provides both visual and mathematical analysis of human's relationships. Social network is also known as web. This paper focuses upon the types of SNA and its uses. The paper also discusses some of the social networking tools by presenting a comparative study between them.

Key Words: Social Networks Analysis (SNA); SNA tools; Gephi; Pajek; Networks; IGraph, nodes, RFID

I. Introduction

Social Network Analysis is a concept measuring relationships and flow between people, groups, organizations, computers & URLs and other connected information or knowledge entities [1].

It is a process of investigating social structures through the use of networks and graphs theories. Its network structures are in the terms of nodes and the ties, edges or links that connect them.

Social network analysis is used in a wide range of applications. Some common network analysis applications are network modeling and sampling, community-maintained resource support, link prediction and entity resolution etc. Social network analysis is also used in intelligence, counter-intelligence and law enforcement activities [2].

It is a network of set of socially relevant nodes connected by one or more relation.

Hanish Chauhan*
Department of Information & Technology,
Institute of Innovation in Technology & Management,
G.G.S.I.P.U, New Delhi

Surbhi Kakar**
Department of Information & Technology,
Institute of Innovation in Technology & Management,
G.G.S.I.P.U, New Delhi

II. Problem Definition

Social Network Analysis is the process which is used to investigate the social structure with the help of network and graphs. It is a structure which is used to communicate with each other. In this paper, the author talks about nodes, networks and graphs. The concept of social network analysis tools is also discussed in this paper for understanding the social network analysis properly. The first section presents the concept of SNA. In the next section, the paper discusses about the types and uses of social network analysis. The types and uses of SNA show the perfect mirror image of the concept of SNA.

The section thereafter, emphasis on Social Network Analysis Tools like Networkx, Gephi, IGraph and Pajek, presenting a comparative study between them.

SNA tools are used in many fields for collecting information, communication etc.

III. Types of Network Analysis

Broadly, the networks can be divided into two types, Ego Network Analysis and Complete Network Analysis [1]: -

a. Ego network analysis- It is considered as analysis of individual nodes. Network may have various egos as nodes in the graph. Egos can be persons,

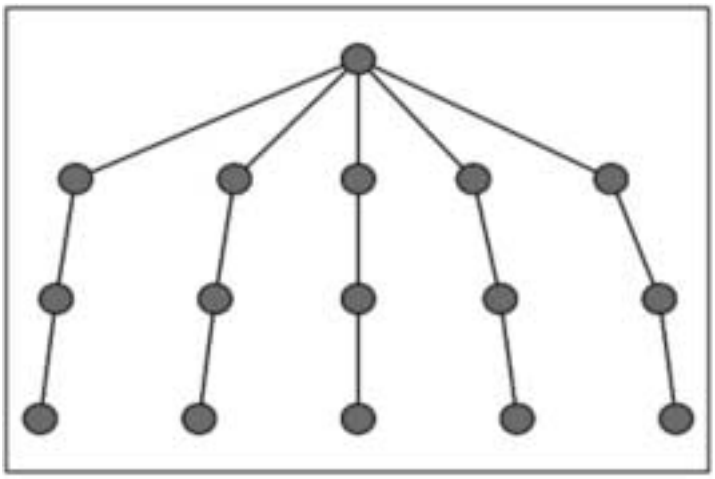


Fig. 1. Networkx (SNA Tool)

groups or whole society. In this network analysis, distinct behavior and its variation is extracted and described.

b. Complete network analysis- It is considered as the analysis of all the interactions among a set of nodes. Techniques such as subgroup analysis, equivalence analysis and measures like centrality closeness, degree, and betweenness all require complete networks.

IV. Uses of Social Network Analysis

Following are some uses of SNA: -

- Identify new technical trends becoming commercially feasible, e.g. RFID, Genome sequencing, tissue engineering. [4]
- Analyze skillful network, Co-authorship networks, concitation networks, patent networks.
- Degree of success.
- Ranking of trends, of authors, of corporations commercializing trend. [4]
- Analyzing page status, Page Rank (Related to recursive in-degree calculation), Specialists/Hubs. [4]

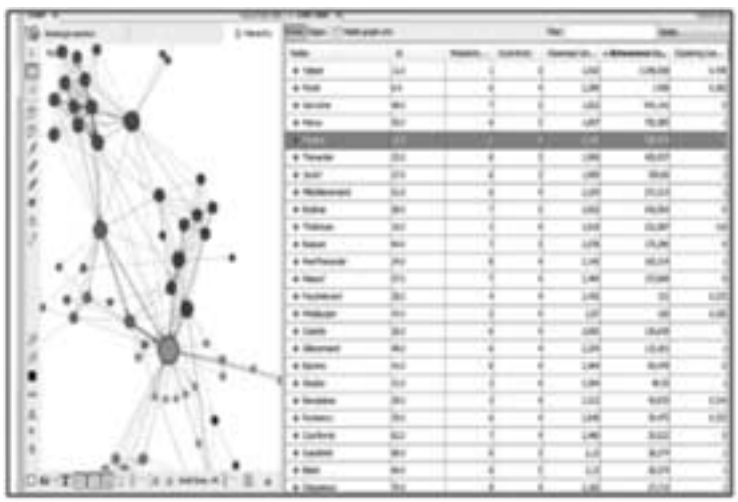


Fig. 2. Gephi (SNA Tool)

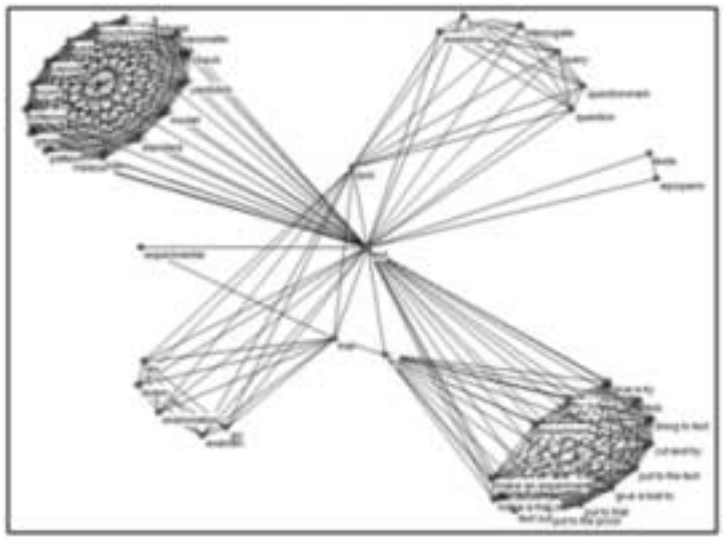


Fig 3. Pajek (SNA Tool)

- Discovering Groups: Finding nearest sets of nodes. [5]
- Analyzing Trust: Propagating Faith.
- Using propagated faith to fight junk mail: In Email, In Web page status. [4]

V. Social Network Analysis Tools

Social network analysis tools are used to identify, analyze, visualize or simulate nodes, from various types of input data including mathematical models of social networks.

The author displays a study on four analysis tools namely Gephi, Networkx, IGraph and Pajek. All these tools are freely available for use. Here is the brief about each of them-

a) Networkx- It is Python language software. It is used for creation, manipulation of a complex network. With this tool one can load and store

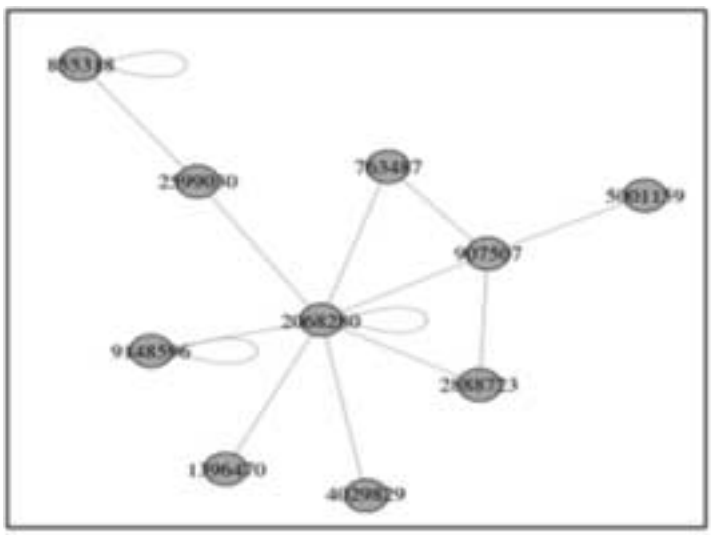


Fig. 4. IGraph (SNA Tool)

networks in standard data formats. Networkx has many features like language data structures for graphs, dIGraphs, and multIGraphs etc [1].

b) Gephi- It is an interactive visualization and Exploration platform. It has dynamic and Hierarchical graphs. It is helpful for those people who explore and understand graphs.

c) Pajek- It is used for Software for drawing networks. It has analytical capabilities and compute centrality measures.

d) IGraph- It is a free software. It is used to create and manipulate graphs. It includes theory problems and network flow. It is also used to handle millions of nodes and edges. It can be installed as libraries for python and ruby [1].

VI. Comparison Between SNA Tools

Table I shows a brief comparison between the tools Gephi, Igraph, NetworkX and Pajek. The comparison is done on the basis of the software, file formats and the graph types used for these tools. [3]

Table I. Comparison Table

Software	Gephi	IGraph	NetworkX	Pajek
Version	0.7 alpha	0.5.3	0.6	1.26
Type	Stand-alone software	Library	Library	Stand-alone software
Platform	Java	Python/ C Libraries	Python	Windows
License	GNU GPL	GNU GPL	BSD License	Free for non-commercial use
Expectable Computing Time	Medium (Java)	Fast (C)	Fast (C, Python)	Fast (C)
Trackable No. Of Nodes	500,000	>1.9 million relations	1,00,000	500,000
Time to load 10^6 nodes & edges	40 sec.	11 sec.	137 sec.	24 sec.
File Format				
GML (Geography Markup Language)	It supports GML format.	It supports GML format.	It supports GML format.	It doesn't support GML format.
GraphML	It supports GraphML format.	It supports GraphML format.	It supports GraphML format.	Export only.
DL (Dick image)	It supports DL format.	It doesn't support DL format.	It doesn't support DL format.	It supports DL format.
GEXF (Graph Exchange XML Format)	It supports GEXF format.	It doesn't support GEXF format.	It doesn't support GEXF format.	It doesn't support GEXF format.
Graph Types:				
Two Nodes graphs	It doesn't support two nodes graphs.	It supports two nodes graphs.	It supports two nodes graphs.	It supports two nodes graphs.
Multi-relational graphs	It doesn't support multi-relational graphs.	It doesn't support multi-relational graphs.	It doesn't support multi-relational graphs	It supports multi-relational graphs
Temporality	It doesn't support temporality.	It doesn't support temporality.	It supports temporality	It supports temporality

VII. Conclusion

Stand alone software is very beneficial for graph Visualization, data conversion. IGraph is efficient tool that offer most of graph features and manage large and difficult network. Libraries (Networkx or IGraph) are more efficient for tasks involving millions of nodes and for operations such as the union and the difference between sets of nodes or for the clustering. Stand alone software are convenient to use to learn, so for beginners Pajek and Gephi is appropriate software.

For complex dataset and research purpose the author can use Networkx and IGraph software.

For one mode or two mode network analysis the author uses any of software tools but for multi-relation network graph, the author has only Pajek software tool & for temporarily network graph the author has Networkx and Pajek tools. All of the above mentioned software can use .Net file format. But typically data are present in .txt format.

IGraph offers generally graph features and it also manage large and difficult network. IGraph and Pajek are more rapidly software Tools compare to others. But, Pajek doesn't offer all graph features. In case, when one wants to analyze all graph features they can use IGraph software. IGraph gives quick results to nearly all graph features. Its load time is less then Pajek software. So based on execution time IGraph is better software.

The future scope of this paper includes categorizing the above tools based upon time, space complexity and the application areas they are used.

References

1. Akhtar, Nadeem. "Social network analysis tools." Communication Systems and Network Technologies (CSNT), 2014 Fourth International Conference on. IEEE, 2014.
2. Sparrow, Malcolm K. "The application of network analysis to criminal intelligence: An assessment of the prospects." Social networks 13.3 (1991).
3. Combe, David, et al. "A comparative study of social network analysis tools." Web Intelligence & Virtual Enterprises. 2010.
4. Nextmedia, C. S. A. "Social Networks Overview: Current Trends and Research Challenges." European Commission Information Society and Media (2010).
5. Business Application of Social Network Analysis BASNA-2013, (2014, Nov 18), [Online], Available: www.basna.in

Big Data Mining in Cloud Environment: Challenges & Future Scope

Harsh Arora*
Deepti Gupta**

Abstract

The enormous growth in amount of data that has been produced over last few years to an exceptional level due to use excessive use of Internet and its applications where everything from machines to devices are loosely connected has made Big Data as one of the most researched topic in IT presently. The volume of unstructured data that is generated at very high velocity and collected from variety of heterogeneous sources like social networking sites, Health surveys, customers transactions, Weather forecasting etc has modified the way data is processed using traditional data mining techniques. Comprehensive data computing and enhanced tools are required to store, process, analyze and visualize the Big Data. Cloud Computing has transformed the way organizations are analyzing data to make profitable decisions by improving availability and reliability but at the same time poses some challenges like degraded speed of WAN in moving large clusters of data into and out of the cloud. This paper discusses the challenges and scope of mining Big Data in cloud environment.

Key Words: Big Data, Data Mining, Cloud Computing

I. Introduction

In today's environment-"Big Data" has become a buzz word which in true sense is a word having millions and trillions of data. In this era, there has been increasing demand of data and its processing due to introduction of number of internet technologies, websites and social network applications like Facebook, Twitter etc. Although users can create data or content freely on the web but tremendous increase in data has amplified the web volume. Now on one side if the users are comfortable in accessing various applications over the web, on the other side it has become a big challenge to handle the big data as it contains every type of data (structured, semi-structured and unstructured). For the matter, cloud concept has been introduced where data is present in aggregated manner at different levels so that different cloud servers can increase the efficiency of data execution including server, OS and middleware for sharing them. Cloud provides a reliable, fault tolerant and scalable environment so that big data systems can perform efficiently.[4] The idea here is to handle and manage bulk of data by using various mining tools and techniques. This ocean of data called big data is created with great velocity hence it must be processed and monitored with high frequency over the cloud. Further security is another parameter that must be taken into account. It is required to integrate privacy or security measures on big data levels to improve the performance of the system. Apache Hadoop which complements MapReduce algorithm has been introduced for Big Data Mining. This platform has been introduced in order to manage, process and integrate the massive amount of complex and unstructured data present over the internet. Such frameworks are based on concepts like distributed storage, massive parallel processing and fault tolerant and scalable solutions. Data on cloud

Harsh Arora*
Institute of Innovation in Technology & Management
Janakpuri, New Delhi

Deepti Gupta**
Institute of Innovation in Technology & Management
Janakpuri, New Delhi

must be encrypted and for that matter it s required to apply various security methods and techniques on it.

II. Mining the Big Data: Importance and Objective

Big data has become a challenge in the market as it is a collection of complex and large datasets that are difficult to process and mine for the patterns and knowledge using traditional data management tools. It is the matter of heterogeneous and diversed cum distributed type of data. Now the question arises on how to organize and control this bulk of data in a form where vulnerabilities are present and which provides ease of accessing data all over. Hence there should be some mechanism to handle such distributed big data in order to provide feasible and easily accessible platform to users all over the world without any complexities. Traditional data mining tools and techniques are not efficient to process Big Data which is the main concern of this paper. Many organizations are increasingly generating data in big amount as per their requirement of business processes ,multiple user activities plus there are social networking sites that are having multiple user records on daily basis like their daily posting details of activities they perform, events attended, places they visit, pictures they take and things they enjoy.[2]. It requires storage, management, interoperability, governance and analysis of data to be considered seriously.

Big data Mining has great importance and objectives as it is the matter of handling, managing , controlling, filtering and processing huge data in systematic and organized manner in order to provide error free and simple and flexible platform to the users to access data. In Data Mining, interesting patterns or knowledge is discovered in data which is basically mining knowledge form data. Moreover knowledge extraction, pattern analysis, data archaeology and data dredging is done in in Big data Mining. The main purpose of data mining is to discover previously known interrelations among apparently unrelated attributes of data sets by applying methods form several areas including machine learning, database systems and statistics. [2] Analytics comprises techniques of KDD, data mining, text mining, statistical and quantitative analysis, explanatory and predictive models, and advanced and interactive visualization to derive decisions and actions.

In true sense, we need new techniques and algorithms to deal with ocean of data" Big Data' that is having different data types (structured, unstructured and semi-structured).The five V's that describe the characteristics of big data are:

V1: Volume (Related with big amount of data available and factors associated that contribute to increase in volume of data in form of petabytes and zettabytes.)

V2: Variety (Related to every kind of data like structured that is relational databases, unstructured including audio, videos, texts and images, semi-structured data in XML format)

V3: Velocity (It describes the rate at which data is growing or the speed of generation of data from multiple heterogeneous sources.)

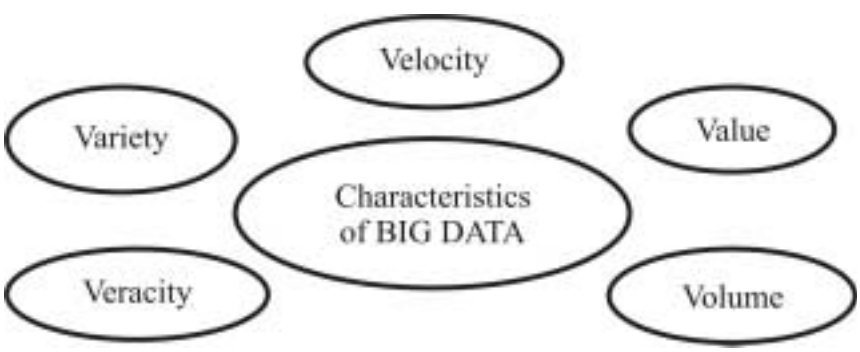


Figure 1: Characteristics of Big Data

V4: Veracity (This defines the quality of data that is being mined to extract knowledge.)

V5: Value (It is related to the monetary worth that organization is going to get on mining the data.)

Big data is having applications in every field these days like in Business, Human behaviour Monitoring, Smarty cities, GIS, Bio- Informatics, medical Science, Weather Forecasting, Cloud control system, multimedia, body sensor networks etc. Now it has been concluded from mining importance objectives that apart from traditional models of big data, we need some more training models which can be used to estimate the parameters to handle and manage the big data.

III. A View of Big Data in Cloud Environment

Cloud computing is a technology which is based on sharing and aggregating the computing resources at one place with reliability, security, availability and scalability so that big data systems could perform in efficient manner. The principle objective of cloud computing is to deliver a platform of computation of big data from multiple sources and providing solution for processing of such enormous amount of information.

A. Big Data processing in cloud environment

Big data and cloud computing are compatible terms in current era because we need a platform that can provide services at one place where sharing , pooling and aggregation can be done easily and on time with great flexibility and reliability. Moreover security is another important feature that is to be taken into account. Hence cloud is a collection of servers and database that has been created with the objective of delivering services like computation of data at a centralized location through internet. So in this case, local computers are no longer need to manage data or applications. Cloud computing uses the mechanism of network of large group of servers for distributed data processing among servers that leads to parallel execution of millions of instructions per second. The advantage of using cloud for Big data mining is that it reduces the overall cost of mining the big data by eliminating the need to purchase personal IT infrastructure.

It is required for the user to run the cloud application to be able to connect with cloud interface and use its services. There are many real time applications which are using cloud computing like Gmail, Dropbox, Google docs etc. In cloud computing, data and software applications are developed and implemented as services. These services are in the form of multi-layerd infrastructure which is Saas that is software as service where applications are hosted and delivered online through web browser offering traditional desktop functionalities like using pixlr.com for editing any image. Another service provided by cloud is Paas that is using some platform available on cloud to execute the code such as using Beanstalk. Third one is Iaas or Infrastructure as service where set of virtual computing resources such as storage and computing capacity is hosted in the cloud. [6]

Further for processing big data over the cloud, data from multiple sources has been accumulated and real time systems are used for parallel complex event processing (CEP). There are also analysis systems present on cloud for processing, aggregation, statistical optimization and parallel data processing in parallel batches efficiently. Resultant values or outputs are displayed through web services on the cloud which can be used for decision making. So cloud computing promises on demand services to users. Cloud ability to virtualize resources allows abstraction of hardware and require little interaction with cloud service providers and enables users to access terabytes of storage, high processing power and high availability in pay as you go model. [4]Hence cloud provides big data accessibility having variety of data. Moreover database handling basic tools are available for structured data, but it is cumbersome to manage unstructured data that has become a challenge when it comes to big data mining. For relational database, SQL is used but NoSql and NewSql has been designed to store unstructured large amount of Big Data. It gives better performance in distributed environment to handle unstructured data.

B. Security issues associated with big data over cloud

On one side where cloud has given the platform for efficient execution, processing and availability, at the same time , cloud comes with security challenge. As data on cloud is present as an open platform that's why we don't require different local or other servers for storage and managing data but owner might not have any control of where data is placed in the cloud. As multiple resources are available on cloud, so it is needed to encrypt or protect data by end to end encryption so that no intruder could interrupt on the network where applications are hosted to be used by users. Data must be integrated and secured on the internet.

Since cloud involves extensive complexity, we believe that rather than providing holistic solution to secure the cloud, it would be ideal to make noteworthy enhancements in security to make the cloud secure.[3] Although with the increase in big data in market due to many web services and applications, many companies are struggling with the security issues. In order to maintain the balance between big data and its security, we need network security techniques and IDS algorithms to protect data from threats. Finally access control framework and third party secure data publication should be applied to cloud to ensure its safety.

IV. Big Data Mining with Apache Hadoop

We are living in an era where data is generating at very rapid rate that it is not possible to use traditional methods of mining to process the data to extract some useful patterns out of it. As data is building block for any organization, the main objective of Big Data Mining techniques is to extract valuable information and knowledge, perform detailed analysis and provide new opportunities and advantages to perform decision making. Due tis very large size and unstructured nature, Big data mining is one of the hottest trends in IT presently. Big data environment requires MAD skills which differs from traditional data warehouse environment. MAD that is Magnetic Agile Deep means big data environment needs to attract data from all sources regardless of their quality, can synchronize with rapid data evolution and should be deep enough so that analysts can perform drill down-up operations to study and analyze such large datasets. Fast data loading, query processing and efficient utilization of storage space are the key requirements of Big data mining.

One of the frameworks that has been developed for Big Data mining is Hadoop that provides complete infrastructure for performing Big Data analytics by providing implementation for MapReduce programming paradigm and Hadoop distributed File System(HDFS) to store such large amount of data nearly in petabytes. HDFS is an optimized distributed file system that provides storage for large files. Data replication is used to ensure reliability and availability of data in case of node failures where data is stored. HDFS is composed of two types of nodes :Name node that directs the client to the Data node where data is actually stored in file blocks. MapReduce programming model was introduced by Google in 2004 and was later adopted by Apache Hadoop. It is based on scaling out rather than scaling up that is adding more resources rather than increasing power of single computer. MapReduce is parallel programming model that performs two functions that is mapping and reducing in parallel using Job tracker and Task tracker in order to reduce the overall time required to complete any task. Job tracker assigns map and reduce function to the available task trackers which will actually run the job.

Map function partitions large unstructured computational task into structured key/value pair which is then used as input to the reduce function which combines all values for a particular key to produce the final output. To explain the working of MapReduce algorithm, let's take an example. We have temperature of four major cities of India-Delhi, Mumbai, Chennai and Kolkata for the month of January and our aim is to find minimum temperature for all the cities. The key would be the name of the city and its temperature would be the value for the corresponding key. Now the Jobtracker will partition this data into four files week wise and then distributes

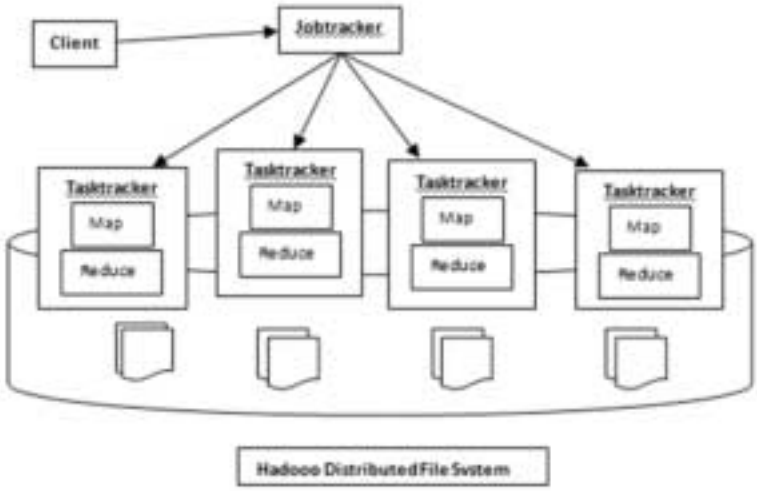


Figure 2 : MapReduce Implementation using HDFS

it across the Tasktrackers. First the Tasktracker will run the mapper and perform map function to find one key value pair for each city per week which is then stored in data node of HDFS.

Suppose the output of four mappers (considering only Delhi) is

(Delhi, 16) (Delhi, 14) (Delhi, 15) (Delhi, 12)

Next the four key value pairs for Delhi (from four mappers) will be reduced to single key value pair combination by running reducer of Tasktracker in order to produce the minimum temperature for Delhi.

Output of reducer: (Delhi, 12)

Same procedure will be repeated for other three cities also. This is how both HDFS and MapReduce work together on Hadoop platform to mine the Big Data.

These results from MapReduce are stored in HDFS and are analyzed and visualized by analysts to perform decision making. Hadoop is designed for offline processing through read instructions rather than relational databases that were used for online transactions making analysis easier for large datasets on Hadoop.

V. Challenges and Future Scope

Although cloud works on pay as you go principle and thus eliminates huge investments in IT infrastructure by providing large storage capacity, still it comes with various challenges to process Big Data that are required to be resolved to provide efficient patterns from mining. Moving terabytes of digital data into and out of the cloud using traditional protocols like FTP, HTTP that introduces unacceptable delays due to degraded transfer speed of WAN is one of the main challenges for Big data computing in cloud environment. Another important challenge is efficient resource allocation to Big Data analytics tasks on cloud. Also the processing of unstructured data is not possible with traditional methods of mining. As data is very valuable to any organization to run their business, efficient safety measures and disaster recovery mechanisms need to be adopted and applied to big data computing environment to ensure continuous availability of data because data loss is equivalent to loss of value. Data governance policies need to be enforced to retain valuable and important data and to manage it throughout data accesses. One of the drawbacks of implementing MapReduce in cloud is that requires customers to learn new set of API's to build solutions for analyzing data. There is a need to provide user friendly tools to visualize the results of mining so that useful and efficient patterns can be extracted and understood. Aspera On Demand by IBM has provided a solution to deliver efficient and scale out transfer capacity to optimize transfer of files

of any format and size over network under any conditions. Quality of Service based data uploading mechanisms may help to ease the process of uploading data on the cloud.

VI. Conclusion

In this research we have analyzed that Big data is very crucial and innovative topic from industry as well as research point of view. Extracting hidden patterns and valuable information from Big data using traditional tools and techniques is not efficient process due to its unstructured nature. Cloud computing definitely provides services at various levels and helps to reduce the overall costs by eliminating the need to purchase personal IT infrastructure by providing resources on demand and scaling infrastructure up and down according to actual demand, still mining big data in cloud environment poses many opportunities and challenges due to its voluminous variety generated at very rapid rate. The slow movement of such a large amount of complex data into and out of the cloud for computing at very degraded speed of WAN and ensuring safety and privacy of valuable data needs continuous attention and development. Although frameworks like Hadoop which provides implementation for Map Reduce programming paradigm and uses HDFS to store large files of unstructured data has been created for mining Big data, still more research and developments are required to find solutions to the problems and overcome the challenges.

References

1. Nada Elgendy, Ahmed Elragal, Big Data Analytics :A Literature Review Paper, Springer International Publishing Switzerland 2014
2. Marcos D. Assuncao, Rodrigo N. Calheiros, Silvia Bianchi, Marco A.S Netto, Rajkumar Buyya, Big Data Computing and clouds: Trends and future directions, Journal of Parallel and Distributed Computing, Elsevier,2015
3. Venkata Narasimha Inukollu, Sailaja Arsi and Srinivasa Rao Ravuri, Security Issues Associated with Big Data in Cloud Computing, International Journal of Network Security & its Applications(IJNSA),Vol.6, No.3,May 2014
4. Pedro Caldeira Neves, Bradley Schmerl, Jorge Bernardino and Javier Camara, Big Data in Cloud Computing:Features and Issues
5. Khalid Adam Ismail Hammad, Mohammad Adam Ibrahim Fakharaldien, Jasni Mohamed Zain, Mazlina Abdul Majid, Big Data Analysis and Storage in Proceedings of the 2015 International Conferenceon Operations Excellence and Service Engineering, Orlando,Florida, USA,September 10-11,2015
6. Irina Neaga, Yuqiuge Hao, A Holistic Analysis of Cloud Based Big Data Mining, International Journal of Knowledge, Innovation and Entrepreneurship, Volume 2 No.2,2014,pp ,56-64
7. Praveen Kumar,Dr Vijay Singh Rathore, Efficient Capabilities of Processing Big Data using Hadoop Map Reduce, International Journal of Advanced Research in Computer and Communications Engineering Vol. 3, Issue 6,June 2014
8. Prof R.A.Fadnavis, Samrudhi Tabhane, Big Data Processing using Hadoop, International Journal of Computer Science and Information Technologies, Vol.6(1),2015, 443-445
9. Getaneh Berie Tarekegn, Yirga Yayeh Munaye, Big Data: Security Issues, Challenges and Future Scope, Internal Journal of Computer Engineering & Technology Volume 7,Issue 4,July-Aug 2016
10. Swati Harinkhere, Nishcol Mishra,Yogendra P S Maravi, Varsha Sharma, Big Data Processing Techniques & Applications: A Technical Review, CSI Communications ,November 2016
11. S.Rama Sree,K. Devi Priya, An Insight of Big Data Analytics Using Hadoop, CSI Communications, November 2016
12. Wei Fan, Albert Bifet, Mining Big Data: Current Status , and Forecast to the Future, SIGKDD Explorations, Volume 14, Issue 2

Development of Vector Processed IHDPS Expert system for Hydrocarbon reserve estimate in earth & yield prediction supported by DMW techniques

Ghanshyam Yadav*
Anant Gupta**

Abstract

As flexibility and agility become key success factors of competitive oil business industries, the ability to support the short term decision & prediction making of oil exploration, scheduling, oil availability and yield prediction becomes a critical issue. This research paper presents a rule-based expert knowledge system i.e. web-based Hybrid expert system, called Intelligent Hybrid Decision & Prediction System (IHDPS) run on the cyber-enabled .NET Expert System Shell (NESS) technology platform to addresses how engineering knowledge can be dynamically represented and efficiently utilized in oil business industries. The rule-based knowledge system, called web-based Intelligent Hybrid Decision & prediction System (IHDPS), is designed and implemented using the rule-based inference, reasoning & decision making approach. The distinctive technical contributions of (IHDPS) focus on three critically integrated elements: (1) a spreadsheet software for interpreting and evaluating performance data, (2) a knowledge rules for time-series pattern recognition, and (3) an embedded application component.

Key Words: web-based Hybrid expert system, cyber-enabled .NET Expert System Shell (NESS), Intelligent Hybrid Decision & prediction System (IHDPS)

I. Introduction

The main goal of our paper presentation is to develop an intelligent hybrid decision & prediction expert system now recognized as ihdps for predicting the yield of oil and availability of oil using cyber-enabled .net expert system shell (ness) technology platform techniques.

Following are representative application areas where ihdps (vector processor) is used are:

1. Long range weather forecasting
2. Petroleum explorations
3. Seismic data analysis
4. Medical diagnosis
5. Ai and expert systems
6. Aerodynamic and space flight simulations, etc.

Technically, our main concern is to predict the yield of oil and availability of oil based on some geo-logical & geo-physical properties with reference to data warehouses of the major hydrocarbon exploration industries.

To overcome the above challenges, this paper designs and develops a web-based Hybrid expert system, called Intelligent Hybrid Decision & prediction System (IHDPS), using cyber-enabled NESS technology. The objective of this research is to address how the engineering knowledge can be dynamically represented and efficiently utilized to support shop floor production control and job dispatching.

Ghanshyam Yadav*
Dept. of Comp. Sc. & Engg.
JEMTECH, JIMS Group, Greater Noida, India

Anant Gupta**
Dept. of Comp. Sc. & Engg.
JEMTECH, JIMS Group, Greater Noida, India

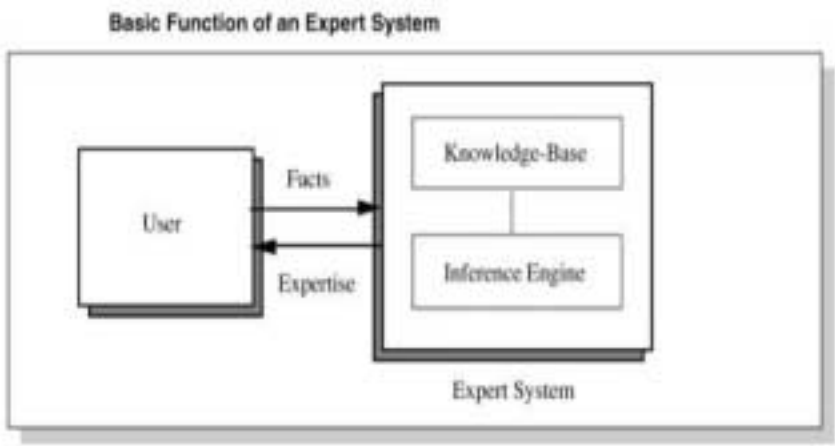


Fig. 1. Functionality of the IHDPS

An IHDPS is A Computer Program Designed To Simulate The Problem-Solving behavior of an Expert in a Narrow Domain or Discipline. "An expert system is a computer system that emulates, or acts in all respects, with the decision & prediction making capabilities of a human expert."

II. Development Strategies Of The IHDPS

Expert systems (ES) are a branch of artificial intelligence to mimic human reasoning and accumulate domain expertise in decision & prediction making. They include a reasoning mechanism called an inference engine that executes a logical sequence of rules and makes calculations to produce answers to a particular problem-solving task. Each type of ES has a corresponding inference technique and knowledge representation approach. It surveys and classifies ES methodologies into the following domains, e.g., rule-based systems, knowledge-based systems, neural networks, fuzzy expert systems, object-oriented methodology, case-based reasoning, and intelligent agent systems.

A. Intelligent decision & prediction system's instruction set architecture

A vector is an ordered set of a one dimensional array of data items. A computer capable of vector processing eliminates the overhead associated with time it takes to fetch and execute the instructions in the program loop. It allows operations to be specified with a single vector instruction of the form

$$C(1:100) = A(1:100) + B(1:100) \quad (1)$$

The vector instruction includes the initial address of the operands, the length of vectors and the operations to be performed, all in one composite instruction. A possible instruction format for a vector instruction is given below. Instruction Format for Vector Processor. This is essentially a 3 address instruction with 3 fields specifying the base address of the operands and an address field that gives the length of the data items in the vectors.

This assumes that vector operands reside in the memory.

Some examples where vector processing is implemented are

- Processing of vector units
- Matrix multiplication
- Memory interleaving
- Super-scalar processors
- Supercomputers

Super-scalar processors, Supercomputers are the vector computers and are characterized by very high execution speed, large main & secondary memory and use of parallel structured software in large extend. These are designed to perform large matrix computations, looping conditions (for, while loop) in the area of hydrocarbon prediction, exploration & artificial intelligence.

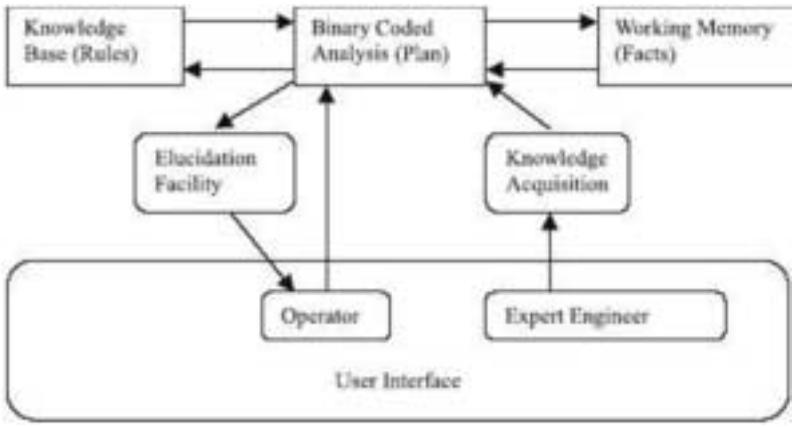


Fig. 2. Rule based Hybrid decision & prediction system with 3- tier architecture

B. In any expert systems development, the most important aspect is to find an expert in that specific domain. The expertise here does not mean just possession of knowledge or qualification but involves problem solving skills that need be sharpened in a particular domain.

Knowledge base-

- Obtainable from books, magazines, knowledgeable persons, etc.
- Stores all relevant information, data, rules, cases, and relationships used by the expert system

Inference engine -

- Draws conclusions from the knowledge base.
- Seeks information and relationships from the knowledge base and provides answers, predictions, and suggestions in the way a human expert would

Rule-

- A conditional statement that links given conditions to actions or outcomes.

Fuzzy logic-

- A specialty research area in computer science that allows shades of gray and does not require everything to be simply yes/no, or true/false.

Backward chaining-

- A method of reasoning that starts with conclusions and works backward to the supporting facts.

Forward chaining-

- A method of reasoning that starts with the facts and works forward to the conclusions

C. Intelligent hybrid decision & prediction systems

Hybrid systems seem to be combinations of fuzzy logic with neural networks. It supports inference chains and neural net training strategies, providing a coherent system for exploiting the advantages of the conventional rule-based expert systems .The pioneering applications of neural nets for implementing expert systems as reported in literature may be classified into the following three categories:

1) Connectionist rule based expert knowledge systems: which are entirely based on neural network technology? The simplest case is to deal with implicit, fuzzy (and possibly shallow) knowledge using a neural net associative memory. In these applications, explicit rules of decision & prediction making are not available or may not be necessary, e.g. machine fault diagnosis etc. The complicated cases could be to acquire rules using neural net structure and make the system capable of explaining the decision & prediction making procedure.

2) Rule oriented neural networks: It incorporate rules into the neural net structure and learning procedure. These applications use rules as a means to inject expertise into neural networks (Nets) and direct the

neural nets to learn and represent the knowledge of interest instead of arbitrary regularity that may be found in the training data. For instance, It use explanation-based learning rule to construct the initial network structure and determine the initial weights of connections to improve learning quality and performance of back propagation networks. Rigorous attempts have been made to design complicated neural network systems for realizing certain types of production rules and inference mechanism.

3) Hybrid inference chains: which treat neural nets and rule as inference elements and use such elements alternatively in the inference process as and when appropriate? In these systems, rules are used for explicit knowledge inference and neural nets for implicit knowledge inference. For instance, neural nets are often used for pre-processing to reduce the noise in the field input, so that rules can be more accurately activated. Also, in some applications a neural net may serve as the central inference engine and rules may be used for pre-processing and post-processing.

D. Categorization

Category 3 includes all the intuitive hybrid systems exploiting the potential benefits of the advanced technology of both neural nets and rule based systems for improving the quality of expert systems in real life applications. Although the examples we cited are all simple static serial combinations of neural nets and rule bases, in fact there is no reason why a hybrid system may not have a multiple level dynamic inference chain which may vary the order of neural and rule inference mechanisms in accordance with the dynamics of the application environment.

Category 2 applications take the advantage of rule based techniques for improving the neural nets performance. In its knowledge acquisition phase, it involves rules and neural net learning algorithms, while in the consultation phase; it is only a neural associative memory.

Category 1 applications attempt to develop some alternative neural framework for implementing expert systems. Since category 2 and 1 applications are application specific and usually involve neural net model development, they may better be viewed as potential neural techniques to be used for constructing hybrid systems like those in category 3.

We may conclude that a hybrid expert system shell should focus on supporting the category 3 applications, provide some mechanisms to support category 2 applications and be easily extensible to support category 1 applications. In addition to the features of a traditional rule based expert system shell the following facilities should be provided:

- Built-in neural net components;
- A way of chaining neural nets with rules;
- A neural net training strategy that can be programmed using rules. Hybrid expert systems with a two level (static) architecture

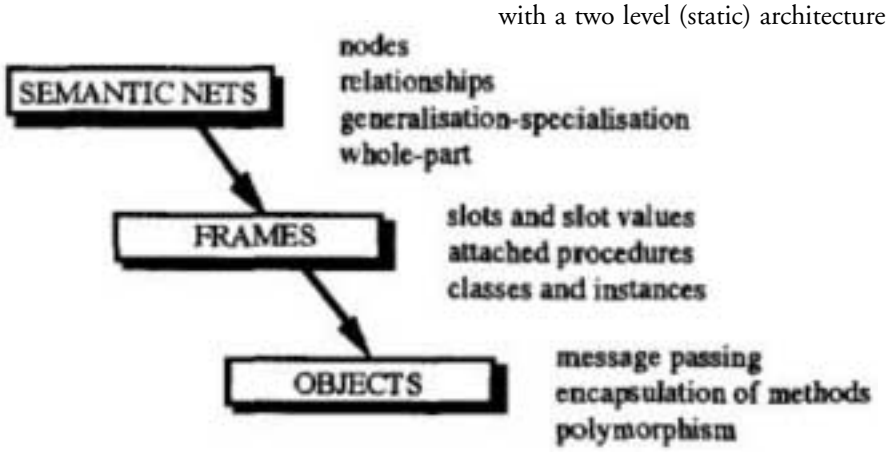


Fig. 3 Knowledge Representation Formalisms

- A lower level neural net for input data pre-processing.
- A higher level rule base has been first introduced as a Rules are **"IF THEN"** statements that have a left-hand side model in. However, as discussed above, a hybrid shell should be able to support arbitrary inference chains.

E. Knowledge representation

Knowledge representation is concerned with the way in which information might be stored and associated in the human brain, usually from a logical perspective (not biological). Knowledge base was extensively indexed and made content addressable, so that any system using it can control the way in which different pieces of knowledge were activated without having to know exactly how they are stored? Sufficient care has to be taken to organize and integrate different pieces of knowledge from the problem solving point of view. The knowledge representation used here is a symbolic representation, which helps in nonnumeric computation in which symbols and symbol structures could be constructed as standing for various concepts and relationship between them.

ES can be broadly categorized into three knowledge representation techniques, i.e., rule-based, case-based, and frame-based. The Rule-based expert system (RBES), also called knowledge- based expert system (KBES), gathers a set of factual data (e.g., program descriptions/degree requirements) and produces actions (e.g., recommendations/schedule) by interpreting the data with a set of predefined rules . RBES needs to collect, extract, and manage adequate domain knowledge constantly.

The main knowledge base components of this system are

- Fact base, which represents the initial state of the problem, and
- Rule base, which contains the operators that can transform the problem state in to a solution.

The inference engine matches these facts against the rules to see which rules are applicable.

It works in a **cyclic** manner as follows:

- MATCH the facts against the rules.
- CHOOSE which rule instantiation to fire.
- EXECUTE the actions associated with this rule.

Part (IF-portion) and a right-hand-side part (THEN portion). Fact based represents the initial state of the problem. It is data from which inferences are derived. When the parameter of fact in the inference engine is changed, the result of inference will also be updated.

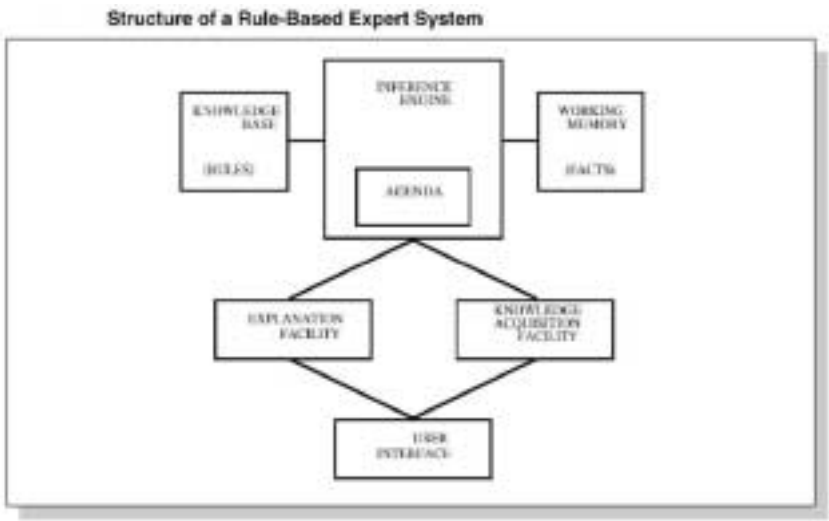


Fig 4: Structure of Rule Based Expert System

The NESS inference engine controls the overall execution. It works in a match-choose-execute cycle. Whenever the rules match with existing facts, it chooses which rules to fire. The execution engine enacts the right-hand-side actions associated with the firing rules.

IHDPS can capture the (senior) expert's know-how on design solutions; they involve the construction and usage of the following modules:

- Knowledge Base (KB)
- Inference Engine
- User's Interface.

This research presents a rule based knowledge system run on the .NET Expert System Shell (**NESS**) platform to addresses how engineering knowledge can be dynamically represented and efficiently utilized in decision & prediction & prediction making. The knowledge system, called Intelligent Hybrid Decision & prediction System (**IHDPS**), is designed and implemented using the rule-based inference and reasoning approach. The distinctive technical contributions of IHDPS focus on three critically integrated elements: -

The distinctive technical contributions of IHDPS focus on three critically integrated elements:

(1) Spreadsheet software for interpreting and evaluating performance data.

(2) Knowledge Based rules for recognizing time-series patterns.

(3) An embedded system application component.

III. Functional Specifications for IHDPS Development

The functional specifications are developed in this stage.

The specification of each system module is defined in detail.

IHDPS consists of four main modules.

The four modules support four roles of participants in the System platform; The IHDPS is designed and developed in the following four stages:

Stage 1 Requirements gathering and system analysis

★ Knowledge management module
★ Rule management module
★ System administration module
★ Project management module Stage 2 Operational scenario designs
★ External Application environment
★ Rule authoring environment
★ Rule execution environment

Embedded application component usage: IHDPS is designed to package its kernel modules and the NESS inference engine as EAC which can be operated independently in any .NET run-time environment and integrated with existing information systems.

The effective use of a KBS is with respect to two main objectives:

★ To support a successful transfer of knowledge from senior expert engineering designers to junior designers.

★ To allow the human expert to use knowledge skills & experience from others within the organization without sacrificing the productivity and design quality.

A. Technical Specifications for IHDPS Development

Data mining classification techniques that can be used in these types of projects:-

1) Decision Tree Induction (ID3,C4.5 & HUNT) Method
2) K-Nearest Neighbor Classifier Approach
3) Support Vector Machine learning Approach
4) Bayesian Classification Approach
5) Case-Based Reasoning Approach
6) Neural network With MATLAB Approach

(1) Neural Network Based Decision Tree Learning Algorithmic Approach for Hydrocarbon Prediction

The results of the two applications are summarized in Table I where the results are the averages generated from 10 fold cross validations for all data sets. Since

Table I: Prediction Results

Measures	C4.5 without Pruning	C4.5 with Pruning	NDT without Pruning	NDT with Pruning
Tree Size (# of nodes)	131	117	67	53
Number of Rules (# of leaves)	52	45	24	17
Test Set Size	320	320	320	320
Correct Classification Rate (%)	86.88	85.31	80.94	79.69
Misclassification Rate (%)	13.12	14.69	19.06	20.31
Mean Absolute Error	0.13	0.15	0.19	0.20
Mean Squared Error	5.51	6.90	11.63	13.20
Computation Time (milliseconds)	15692	14070	28211	26468

the sampling used in cross validation is randomly taken for each run, the results listed include only the ones with the best observed classification rates. From Table 1, it can be seen that in comparing the results generated by the C4.5 decision tree learning algorithm with the NDT model, the NDT model reduces the tree size and number of rules by half with only 5% decrease in classification accuracy. A decrease in the number of rules is an improvement because they are more easily validated by petroleum engineers. The NDT model proposed in this study provides better explanation capability since it generates a comparative smaller rule set with an acceptable level of classification accuracy. The NDT model is able to provide some explicit heuristics for classification with the give up of a little classification accuracy that support predicting oil production from a new well.

One advantage of an inference network in Mineral & ore exploration is explained here: - consider that the in PROSPECTOR, the explicit nature of the network, which makes it easy to identify hypotheses that are impossible or difficult to achieve. If there are missing rules, the inference net may imply the need for these rules. The next uncertainty that is fusion of this evidence how so many different sources are much more difficult sources are much more difficult than evidence in one domain, such as geology.

$$split\ info(X) = -\sum_{i=1}^{n} \frac{|Ti|}{|T|} * \log_2(\frac{Ti}{T}) \quad (2)$$

The condition on which C4.5 selects the test to partition the set of available cases is defined as:

$$gain\ ratio(X) = \frac{gain(X)}{split\ info\ (X)} \quad (3)$$

C4.5 selects the test that maximizes gain ratio value.

(2) Bayesian Classification approach for Hydrocarbon Prediction

I. The BN model consists of nodes representing random variables and edges indicating the interaction between node variables. In our case, the node variables have a geological interpretation, related to prospects and trap, reservoir, or source attributes at these prospects. We use directed edges between nodes, and the resulting graph is known as a directed acyclic graph (DAG). The conditional independence structure imposed by edges simplifies the model specification and introduces sparseness or parsimony. The joint model for a DAG is defined by the marginal distributions for the parent nodes and the conditional probabilities for the children. The joint model for variables ($X_1 \ldots X_n$) is then

$$p(x_1,\ldots,x_n)=\prod_{i=1}^{n}p(x_i|x_i^{pa}), \quad (4)$$

where xpa i denotes the variables at the parent nodes of node i, and this parent set is empty for the top nodes of the DAG. The full conditional distribution of xi, given the outcomes at all other nodes, xi¼(x1,…,xi1,xiþ1, …,xn), depends only on the children, the parents, and the other parents of its children. This is formalized through the notion of cliques and used in the computation of DAGs. The updating of node probabilities relies on the concepts of marginalization and the use of conditional independence structure.

II. Bayesian decision theory under conditions of uncertainty, let's look at the problem of crude oil exploration. Initially the IHDPS must decide what the chances for finding crude oil. If there is no evidence either for or against oil, the IHDPS may assign the subjective prior probabilities for finding of crude oil. With no evidence, an assignment of probabilities that

IV (A). Knowledge Based INference Engine Assembled in IHDPS

Table II: Geophysical and Geochemical Properties

Geo-Physical Parameter	Geo-Logical Parameter
Temperature	Permeability (K) & Porosity (Φ)
Reservoir pressure	PVT of Hydrocarbon holding rocks
Rel. Size & Density	Mineral Composition
water saturation (Sw)	Sedimentary Structure
Dip,(angle of inclination of the reservoir rock's bedding planes)	Back Pressure in well (Surface or bottom hole chokes)
Hade of Structure	Bulk Density (ρb)
Resistivity R	Texture & Lithology
Spontaneous Potential SP	p-velocity in various Rocks
Radioactivity γ	s-velocity in various Rocks
Sonic travel time Δt	p-impedence in various Rocks
Thickness of Reservoir Well	p-impedence in various Rocks
Diameter of Reservoir Well	Degree of cementation
Differential pressure b/wReservoir & Well	Thickness Of The Oil-Bearing Strata
Differential pressure b/wReservoir & Well	Oil saturation (So) & Water saturation (Sw)
Heat (Temp.) w.r.t. time	Formation permeability
Avg. well spacing	Fluid Density & Oil Viscosity
Texture Composition	Rock / pore vol. compressibility
Area & Structure	Rel. Density/Specific Density
Reservoir Mechanics	Capillary Pressure
Formation/Working pressure in reservoir well	Rel. Gravity /Specific Gravity in various Rocks
Casing & Tubing Pressure	Morphologic structure of rocks

are equally weighted between possible outcomes is said to be made in desperation. The term *in desperation* mean that the IHDPS is (necessarily) in desperate need, it is a technical term for the unbiased prior assignment of probabilities. The PROSPECTOR believes that there is better than a 50-50 chance of finding oil.

Using the prior and conditional probabilities, we can construct the initial probability tree. Complete Bayesian decision tree is constructed thereafter for crude oil exploration using backward induction methods.

IV (B). Results & Discussions

Cyber-enabled NESS technology along with data mining techniques was applied to analyze and document the decision & prediction making processes of rules based expert system (**IHDPS**). The monitoring and diagnosis task models were selected from the **NESS** library of generic task models. The validation and planning task models were based on generic task models. These task models were successfully used to facilitate the knowledge acquisition process. They show how procedural knowledge uses domain-specific knowledge to solve problems.

The source reservoir rocks (Crude Oil carrying rocks), their nearby Stratigraphic surfaces and surrounding environment conditions implies serious effects in finding crude oil at any particular places there are basically have following two properties (geo-physical & geo-chemical properties) which are shown in table 2. It forms the knowledge base which is used for further analysis by inference engine that is assembled in IHDPS. It is explained in the adjacent table which is indicated in table 2. Lists of various data mining algorithms that can be applied for hydrocarbon explorations are already briefed in earlier section. A comparative study of decision tree induction approach (ID3, C4.5 & HUNT data mining algorithm) is also shown in table 1. Table 1 shows the prediction results of various decision tree induction approaches (ID3, C4.5 & HUNT).

During the development of the prototype, a few critical points relating to the development of the system emerged. Therefore, to check the validity of the prognosis, some trial data were entered for which the prognosis was already known. Together with the person, two techniques were developed to fine tune and test the method of computing:

1) Variation of the weight assigned to the various goals by means of a graphic interface which enables the user to display a graph showing the weights and to modify them with the mouse.
2) Analysis over time of the prognosis, progressively moving the period of analysis and observing whether the prognosis calculated by the system changes significantly.

V. Conclusions & Future Perspective

The analysis of user requirements is an important and often overlooked issue in expert system development.

The developed system shows the advantages of using Object-orientated analysis and design for the development of expert systems:-

- ★ The user interface can be completely decoupled from the expert system itself because of information hiding and encapsulation.
- ★ The object-orientated approach focuses on developing a model which reflects the problem domain itself rather than the processes present in the problem domain as structured methods.
- ★ The system itself becomes highly reusable. This includes the classes defined during analysis, design and implementation.

A program written in CLIPS may consist of rules, facts, and objects. The inference engine decides which rules should be executed and when. A rule-based expert system written in CLIPS is a data-driven program where the facts, and objects if desired, are the data that stimulate execution via the inference engine. Considerable analysis of their domain and computing background enabled user requirements to be formulated. Ways in which they could be implemented will be explored at the design stage. In conclusion, IHDPS allowed the analyst to cope with a large amount of information and determine user needs whilst also analyzing organizational and domain requirements.

Acknowledgment

We would like to thank to our chairman of JIMS Group & Dr. A. K. Chaturvedi, Director (BBA/BCA, JIMS Gr. Noida) for their time and efforts as our experts guide for research.

We would like to acknowledge my colleagues, faculty members and other contributors for developing and maintaining the IEEE LaTeX style files which have been used in the preparation of this template.

References

1. "Object-Orientated Development of Expert Systems" Dr D S Mehandjiska, email D.Mehandj@massey.ac.nz David Page, email D.C.Page@massey.ac.nz Dept. of Computer Science, Massey University, Palmerston North, New Zealand
2. "Problems in Expert Systems Development" Elizabeth A Kemp (E.Kemp @massey .ac.nz) Massey University, Private Bag, Palmerston North, New ZealandS.
3. "NCLIPS - A Platform for Implementing Hybrid Expert Systems C. J. Wang, E. W. Khor and S. K. Tang
4. "Neural Networks Approach to Rule Extraction" Masumi Ishikawa et. al [12]. Dept. of Control Engineering & Science, Kyushu Institute of Technology
5. HERMES: "an Expert System for the Prognosis of Hepatic Diseases" Bonfh I.a, Maioli **C.b,** Sarti **F.C,** Milandri **G.L.C,** Dal Monte **P.R.**C Computer Systems Sector USL 29 Bologna Est, Via Castiglione 29, Bologna, Italy
6. https://www.researchgate.net/publication/266670693
7. "Strategies for Petroleum Exploration on the Basis of Bayesian Networks: A Case Study" Gabriele Martinelli, Jo Eidsvik, and Ketil Hokstad, Norwegian University of Science and Technology, and Ragnar Hauge, Norwegian Computing Center

Agile Methodology for Software Development

Charul Nigam*
Shaifali Gupta**

Abstract

Agile method for software development is not just any other method to develop software. This Agile Methodology has various processes and approaches to be implemented to develop software. Processes like Extreme Programming (XP), SCRUM, Feature Driven Development (FDD), Crystal Methods, Dynamic Systems Development Method (DSDM), Adaptive Software Development (ASD) and Lean Development are the key implementations of Agile Method of software development. Agility for any software is all about how quick and adaptive the software is in its development phase. Traditional software building life cycles are rigid and are inflexible to the ever changing needs of the business. Agile methods believe in the short and quick deliverable tasks that can give the response or feedback rather quickly than the traditional methods. The risks and defects in the traditional methods are addressed at the end when almost the complete software is ready which takes in a lot of effort and cost to rectify it. While in the Agile Methods, after every short deliverable task, the errors and defects can be rectified in the next iteration or in the next incremental version of the software. Therefore, Agile Approach is also effective in its economic approach wherein the cost factor is always handled. Time management is also seen in Agile Methodology as every deliverable is reached to product owner with in maximum 1-4 weeks. Therefore agile method along with its principal, characteristics and manifesto brought a new technology in which the software is build up to its best efficiency and potential.

Key Words: Extreme Programming (XP); SCRUM; Feature Driven Development (FDD); Crystal Methods.

I. Introduction

Software Development Life Cycle or SDLC for short is the process or structured sequence of stages that is required to build any software. Many organizations make use of SDLC Models to build their software. Any SDLC aims to produce high quality software that meets or exceeds customer expectations, reaches completion within times and cost estimates. All SDLC models have the following stages that must be followed to reach at the completion of the software, these are:

Charul Nigam*
Department of IT
Institute of Innovation in Technology & Management
New Delhi

Shaifali Gupta**
Student - BCA
Institute of Innovation in Technology and Management
New Delhi

Communication: The very first step taken by the customer where he/she initiates the need and requirement of the software to the service providing organization.

Requirement Gathering: After that the software development team looks upon the demand of the customer and try to gather as much information as they about the requirement needed by the customer. This can be done in various ways.

Feasibility Study: To check whether the project or the software they are about to take up is technically, practically and financially feasible or not. Then the process is moved further.

System Analysis: At this stage the team analyses the system for its limitations and impact on the organization as well as on the personnel. The team finds out the limitations of the existing system. It also

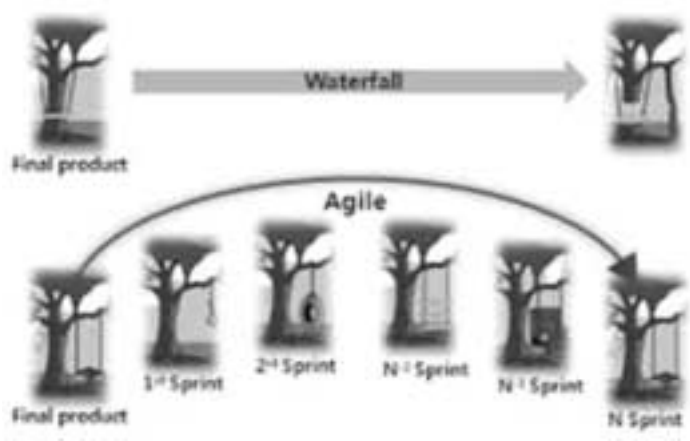


Fig 1.Agile Model against traditional software development method.

then plans the scope of the project and the resources required by it.

Software Design: This design has two parts i.e., logical design consisting of metadata and flowcharts etc. and physical design comprising of the actual look of the software.

Coding: This is the programming phase in which the actual logic is implemented and the code is written in one of the programming language.

Testing: The complete should be tested using different testing criteria like structural testing, function testing, unit testing, system testing, alpha testing, beta testing, stress testing etc. Testing is one of the crucial phases of SDLC.

Integration: Integration means joining of the different modules of the software with one another and with the outer resources.

Implementation: Installation of the software on the other computers of the end users to check portability and adaptability.

Operation and Maintenance: In this phase, the team checks the operation of the software on the end users systems. Maintenance is done over a time period to update any errors if found.

Disposition: When the software is no more of use or is very old to adapt the new technology then it either becomes completely obsolete or need urgent up gradation.

All of these stages or phases are present in every SDLC Model. Some the traditional models are:

- Waterfall Model
- Iterative Model
- Spiral Model
- V-Model
- Big Bang Model
- Prototyping Model
- Incremental Model

But what agile model has to offer is far more than what we have been till now using. Agile model is a combination of iterative and incremental model. What happens in agile model is that when the customer demands for software, the agile team takes in the first and the most basic requirement from the customer of what he/she wants. Then this piece of the task is further broken into smaller pieces known as sprints. In agile methodology, the big software is broken down into tasks which further segregate into sprints. The customer prioritizes the tasks and then every task is done in that priority order. For every task several sprints are created and the team work on sprints. The one by one completion of sprints leads to the completion of one task. When one task is done, it is delivered to the customer and the customer starts using it. Then he/she checks for any loopholes left and reports it back to the team. The team then takes care of the defects in the next iteration of the next task. Like this slowly all the tasks in priority order are

completed with corrections of the defects in next iteration leading to the completion of the software. In this way neither the customer has to wait long for the working software nor has the team to wait for the errors to occur at the end of the completion of the software.

This paper tells about the approaches used in the agile methodology and its various advantages and limitations of the Agile Model.

II. Agile Model Principles And Manifesto

Traditional methods for software development, as discussed earlier, are inflexible to the changes in the business environment. Some of the traditional methods might be cheaper and easy to develop but in the long run in an organization the software must be flexible enough to the demands of the business. Agile methods are nothing but the ever ready nature of welcoming the changes even in the later stages of development. So it is not only about the lines of codes, or the time taken or the cost of the software. It is mainly about the flexibility of the software which is provided by the agile methodology and not by the traditional methods. Traditional methods are considered to be heavy-weighted and documentation driven while agile methods are the alternative to these methods and are considered being "light-weighted".

The term Agile was introduced in 2001. A group of developers independently developed methods and practices to act to the changes they were experiencing in software processing and development. Since then, the Agile Movement, with all its values, principles, methods, practices and tools has significantly changed the landscape of the modern software engineering and commercial software development. The methodologies originally embraced by the Agile Alliance were Adaptive Software Development (ASD), Crystal Methods, Dynamic Systems Development Method (DSDM), Extreme Programming (XP), Feature Driven Development (FDD) and SCRUM.

The manifesto reads as follows (Agile Alliance, 2001):

"We are uncovering better ways of developing software by doing it and helping others to do it. Through this work we have come to value:

- Individuals and interactions over processes and tools.
- Working software over comprehensive documentation.
- Customer collaboration over contract negotiation.
- Responding to change over following a plan
- That is, while there is value in the items on the right, we value the items on the left more."

The previous five values have been further defined by twelve principles:

- Our highest priority is to satisfy the customer through early and continuous delivery of valuable software.
- Welcome changing requirements, even late in development. Agile processes tackle change for the customer's competitive advantage.
- Deliver working software frequently, from a couple of weeks to a couple of months, with a preference to the shorter timescale.
- Business people and developers must work together daily throughout the project.
- Build projects around motivated individuals. Give them the environment and support they need, and trust them to get the job done.
- The most efficient and effective method of conveying information to and within a development team is face-to-face conversation.
- Working software is the primary measure of progress.
- Agile processes promote sustainable development. The sponsors, developers, and users should be able to maintain a constant pace indefinitely.
- Continuous attention to technical excellence and good design enhances agility.
- Simplicity—the art of maximizing the amount of work not done—is essential.
- The best architectures, requirements, and designs emerge from self-organizing teams.
- At regular intervals, the team reflects on how to become more effective, then tunes and adjusts its behavior accordingly.

III. Characteristics of Agile Methodology

Here are some of the characteristics of the agile methodology which proves that why agile alliance came up with such a technique to develop software and why is it so productive and beneficial then the traditional methods of development. The characteristics are as follows:

Modularity: Modularity is a key element of any good process. Modularity allows a process to be broken into components called activities. A software development process prescribes a set of activities capable of transforming the vision of the software system into reality.

Iterative: Agile software processes acknowledge that we get things wrong before we get them right. Therefore, they focus on short cycles. Within each cycle, a certain set of activities is completed. These cycles will be started and completed in a matter of weeks. However, a single cycle (called iteration) will probably not be enough to get the element 100% correct. Therefore, the short cycle is repeated many times to refine the deliverables.

Time-Bound: Iterations become the perfect unit for planning the software development project. We can set time limits (between one and six weeks is normal) on each iteration and schedule them accordingly. Chances are we will not (unless the process contains very few activities) schedule all of the activities of our process in a single iteration. Instead, we will only attempt those activities necessary to achieve the goals set out at the beginning of the iteration. Functionality may be reduced or activities may be rescheduled if they cannot be completed within the allotted time period.

Parsimony: Agile processes are more than just a traditional software development process with some time constraints. Attempting to create impossible deadlines under a process not suited for rapid delivery puts the onus on the software developers. This leads to burnout and poor quality. Instead, agile software processes focus on parsimony. That is, they require a minimal number of activities necessary to mitigate risks and achieve their goals.

Adaptive: During an iteration, new risks may be exposed which require some activities that were not planned. The agile process adapts the process to attack these new found risks. If the goal cannot be achieved using the activities planned during the iteration, new activities can be added to allow the goal to be reached.

Incremental: An agile process does not try to build the entire system at once. Instead, it partitions the nontrivial system into increments which may be developed in parallel, at different times, and at different rates. We unit test each increment independently. When an increment is completed and tested, it is integrated into the system.

Convergent: Convergence states that we are actively attacking all of the risks worth attacking. As a result, the system becomes closer to the reality that we tend to seek with next iteration. As risks are being proactively attacked, the system is being delivered in increments.

People-Oriented: Agile processes favor people over process and technology. They evolve through adaptation in an organic manner. Developers that are empowered raise their productivity, quality, and performance. After all, they are the best individuals in the organization to know how to make these changes.

Collaborative: Agile processes foster communication among team members. Communication is a vital part of any software development project. When a project is developed in pieces, understanding how the pieces fit together is vital to creating the finished product. There is more to integration than simple communication. Quickly integrating a large project while increments are being developed in parallel requires collaboration.

IV. Approachs in Agile Methodology

Agile methods and approaches are not a single process or a single criterion to be followed. It involves various approaches, all different in design and practice. Every approach has its own pros and cons. Each method is totally different but still similar. Every method has its own major principle of the agile methodology.

The main and most commonly used methods of Agile Methodology are:

- Extreme programming (XP)
- SCRUM
- Feature driven development (FDD)
- Crystal methods
- Dynamic systems development methods (DSDM)
- Adaptive software development (ASD)
- Lean development

A. Extreme programming (xp)

It is a methodology for creating software within a very unstable environment. It allows flexibility within the modeling process. The main goal of XP is to lower the cost of change in software requirements. With traditional system development methodologies, like the Waterfall Methodology, the requirements for the system are determined and often "frozen" at the beginning of the development project. This means that the cost of changing the requirements at a later stage in the project - something that is very common in the real-world can be very high. It is one of the first agile processes that have been proposed. In general, XP consists of a set of individual practices that when put together yield a successful software Practice. Further the focus of XP is on the business aspect of a project resulting in increased productivity.

Below are the core practices and values followed by Extreme Programming:

- Whole Team: The team is considered to be very important in XP. The team may consist of developers, who develop the software, testers who are responsible for providing Quality Assurance, and analysts who help in design and the customer representative who provides the feedback. The customer representative may be the actual end user of the system. All members on an XP team use common names and descriptions to guide development and communicate on common terms. This is sometimes known as the metaphor in the terms of extreme programming.
- Planning: The desired features of the software, which are communicated by the customer, are combined with cost estimates provided by the programmers to determine what the most important factors of the software are. This stage is sometimes called the Planning Game. Planning is needed for calculation of estimate of how much effort and cost is needed to develop project.

Fig 2. Extreme Programming practices.

These planning are very effective because the product is visible all the time. There are two types of planning in XP methodology:

a) Release Planning: The customer presents features that are expected by him in the software to developer. The developer reviews these features and find out the difficulties constrained in it. Based on all these constraints and technical challenges an initial release is decided.

b) Iteration Planning: Individual story iterations have shorter time spans which span over a few weeks. The customer presents the features which need to be developed over the next iteration. Based on the features presented, the team estimates the time and cost that may be involved in it. Also each iteration helps in learning about the product.

- Small Releases: The software is developed in small stages that are updated frequently, typically every two weeks. The customer defines the features that have to be in project, represented as stories, which need to be developed. Every story represents the smallest increment to which new features of the system can be added, which usually takes only a few weeks to be developed. The team releases the running and tested software to the customer after every iteration. The customer evaluates the software or releases of the product. The release cycles are very frequent and the customers are always presented first with a software release often. The releases which are delivered frequently undergo continuous integration and thorough testing.

- Simple Design: The design shows the functionality of the system. The software should include only the code that is necessary to achieve the desired results communicated by the customer at each stage in the process. The emphasis is not on building for future versions of the product. To achieve simple design; XP puts an emphasis on using refactoring techniques such as removing duplicated code, improving the existing design. Programmers must verify that the system is still operational after refactoring activity takes place. The XP process requires that all the phases of software development viz. design, implementation, and testing of the system should be carried out by a pair of programmers sharing one computer. This helps programmers to spend more time on finding solutions to challenging problem and less time doing routine debugging. XP programmers improve the design of the software through every stage of development instead of waiting until the end of the development and going back to correct flaws.

- Pair Programming: All code is written by a pair of programmers working at the same machine. This ensures that the programs are reviewed by at least another programmer. This will lead to a better design, testing and code. Also pair programming helps in spreading knowledge of the product across the unit. The pair programmers often switch their work and this leads to all the members in the team participating in pair programming. This helps in raising their skills and understanding of the system.

- Collective Code Ownership: Every line of code belongs to every programmer working on the project, so there are no issues of proprietary authorship to slow the project down. Code is changed when it needs to be changed without delay. In XP, the code is owned by the entire team. No "module" is owned by a single person. In XP, any developer can work on the code base at any given time. This means that the code is seen by many people and this increases quality and reduces defects. If code is owned by an individual, the probabilities of errors or mistrusts are greater. The programmers all write code in the same way. This allows them to work in pairs and to share ownership of the code.

- Sustainable Pace: The team develops the project at constant pace. The focus is on pace that can be sustained indefinitely. Work is carried out in a way such that productivity lasts. Working overtime means more pressure on the members and does not guarantee quality code. The XP team integrates and builds the software system multiple times per day to keep all the programmers at the same stage

Fig 3. Scrum approach

of the development process at once. To keep the development team motivated, XP discourages team members from working more than 40 hours a week. In addition, overtime weeks are usually limited to no more than two weeks in a row. The XP team does not work excessive overtime to ensure that the team remains well-rested, alert and effective.

- Customer Tests: The XP customer checks the software by stating automated acceptance tests. The XP project is directed by the customer who is available all the time to answer questions, set priorities and determine requirements of the project. This helps in validating the features of the product. The team treats the customer tests in the same way as they do programming tests. The story is not successfully implemented until it passes the acceptance test, which is written by the customer. The advantage of this method is to speed up the development process such that if programmers detect a fault in the code they have the right to fix it. A coding standard is specified globally, to make sure that the development team use the same design and coding conventions.

B. Scrum

Scrum is another light weight method victimize for the development of software. Its principle lies in the fact that small teams working cross functionally produce good results.Scrum is more revenue centric with attention on improving revenue and quality of the software. Since being lightweight it can adapt to changing requirements and releases the software in small release cycles called sprints. This method has a framework which needs to be followed during development. The team can choose the amount of work, staff and how to get the work done. This ensures to give the scum team great flexibility and gives a productive work place.

Three important things in scrum are the product owner, scrum master and team.

The product owner specifies the various features of software, the release date and priorities. In Scrum, the Product Owner is responsible for communicating the vision of the product to the development team. He or she must also represent the customer's interests through requirements and prioritization. Because the Product Owner has the most authority of the three roles, it's also the role with the most responsibility.

The scrum master makes sure that the team is functioning properly, productively and enables cooperation across all roles and functionality. The Scrum Master acts as a facilitator for the Product Owner and the team. The Scrum Master does not manage the team. Instead, he or she works to remove

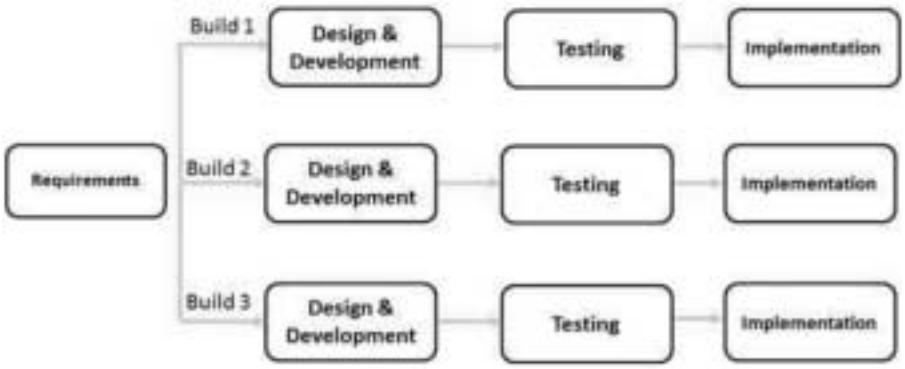


Fig 4. Feature driven development in agile methodology.

any impediments that are obstructing the team from achieving its sprint goals. In short, this role helps the team remain creative and productive. In addition, Scrum has a set of ceremonies associated with it. They include the sprint planning meeting, Daily Scrum Meeting and Sprint review meeting.

- The sprint planning meeting is between the customer and the team. An artifact called the Product Backlog prepared by the product owner has a list of features of the product including functionality and technical architecture.
- The Daily Scrum meeting is a fifteen minute session initiated by the scrum master. The meeting reviews the work that is done regarding development.
- The sprint review meeting held with the customer to discuss the code developed over the last sprint or release cycle. All the stakeholders involved with the product can participate in the meeting and provide inputs for the next sprint. This meeting makes use of two artifacts called the Sprint Backlog and Burn down Chart which record the activities involved in the sprints

Team Member: In the Scrum methodology, the team is responsible for completing work. Ideally, teams consist of seven cross-functional members, plus or minus two individuals. For software projects, a typical team includes a mix of software engineers, architects, programmers, analysts, QA experts, testers, and UI designers. Each sprint, the team is responsible for determining how it will accomplish the work to be completed. This grants teams a great deal of autonomy, but, similar to the Product Owner's situation, that freedom is accompanied by a responsibility to meet the goals of the sprint.

The sprint Backlog is a subset of the product backlog. The sprints or iterations are around 30 days in length. A sprint begins with creating a Sprint Backlog. The team works on the backlog and resolves all the problems mentioned in the backlog. The sprint ends with the system being demonstrated to all the stakeholders. Some Scrum teams have recorded a 10-20% increase in productivity by employing the practices of Scrum. Also progress is made even when the requirements are not stable. Scrum is more of a management process rather than a method meant for developing software. Every iteration is called a sprint which is of duration 30 months. The output of each sprint acts as the input for the next sprint.

C. Feature Driven Development (FDD)

The Feature-Driven Development (FDD) approach focuses on the software features of the system. They are the main driver of the entire development process. It differs significantly from the other agile processes because they put a strong emphasis on planning and upfront design.

The first step of the FDD process is to build a detailed model of the system, which captures all the stakeholders' assumptions and requirements.

Once the domain model is built, the team members print a list of the features of the system. Each feature

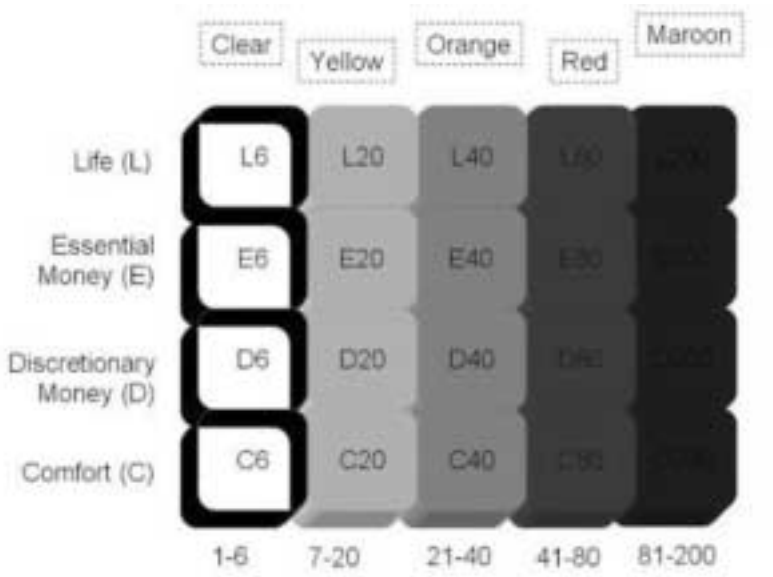


Fig 5. Crystal Methods

has to be developed in a few hours or days, but no longer than 2 weeks.

Using FDD, development teams are divided according to design and to implement a particular feature. The development work is performed in parallel on all the features. Each team is headed by a feature owner, who is responsible for the code segment that implements those features.

This is contrasted with the XP approach where the ownership of the code belongs to the whole development team and not to a specific member.

The FDD process enforces rigorous guidelines in order to find defects in the system. It also enforces coding standards and encourages regular builds on a daily or weekly basis in order to add freshly designed features to the base system.

Since the features are developed in parallel, it is mandatory to have a configuration management system that allows calmly integrating of the changes that are made to the system.

In FDD approach there is a tracking and papering mechanism that is used to show the project status based on the number of features that have been implemented as well as the overall progress of the design, coding, and testing activities.

Each feature is scored using a value ranging between 0 (for a feature that has not yet been worked on) and 1 (a completed feature) and anything in between refers to a feature in progress.

D. Crystal Methods

The Crystal methodology is one of the most lightweight, adaptable approaches to software development. Crystal is actually comprised of a family of agile methodologies such as Crystal Clear, Crystal Yellow, Crystal Orange and others, whose unique characteristics are driven by several factors such as team size, system criticality, and project priorities.

This addresses the realization that each project may require a slightly tailored set of policies, practices, and processes in order to meet the project's unique characteristics.

Several of the key tenets of Crystal include teamwork, communication, and simplicity, as well as reflection to frequently adjust and improve the process.

Like other agile methodologies, Crystal promotes early, frequent delivery of working software, high user involvement, adaptability, and the removal of bureaucracy or distractions.

V. Merits of Agile Methodology

- Improved communication and coordination among team members. Specifically, the daily scrums were seen as a boon and were especially useful to bring testers and developers together.
- Quick Releases is consequence of Continuous Integration, Developers release the sprints of product back log every few weeks instead of every few months or years.
- Developers noted that short sprints having customer feedback led to better agility and efficiency in responding to changing requirements and flushed out bad designs more quickly.
- Most popular benefit is a More Reasonable Process. Many developers complained about rigid development processes that were relaxed in an agile environment.
- The process supports real-time tracking of progress and ability to adjust future forecasts based on real data. Agile methodologies are more dynamic and incur less overhead.

VI. Demerits of Agile Methodology

- Developers feared that the agile process could bring their deficiencies to light. To address this challenge, developers need an environment where they feel safe to expose their weaknesses. If they don't feel comfortable in an open forum then there can be separate meetings for junior or new staff.
- Agile environments seem to blur the boundaries among developers' roles and require competence in a broad range of skills and opposes to the specialization in one. To address this challenge, organizations must strike a balance between team members becoming "masters of all" or "masters of none." It is to be advisable that developers must have broad knowledge on all aspects of software development but should also master their skills in certain areas.
- Agile practices such as pair programming, collocation, meetings, and retrospectives increase social interaction, communication, and the presentation skills.In most of the cases there are people who were technically very talented but have weak communication and presentation skills, while all managers saw the benefits of constant face-to-face communication, the degree of communication in agile environment. Social-skills training are an obvious solution to this challenge. It can help new people to get into the company values easily.
- Some projects can or some can't implement agile values and principles. Some implemented agile methods "on paper," but they didn't achieve agility's ultimate goals. It may be due to many reasons like intangible combination of staff personality, management style and other factors. Formal training is a typical solution to teach agile practices.Some companies included a provision for training and attendance at agile conferences focusing on its values and principles. Continuous training was preferable to one-off training in helping developers absorb and retain agile values and principles.
- It can work for small co-located teams, but not for complex large projects. It can be difficult for larger teams to be as flexible as smaller teams with respect to design and architectural changes. Scrum meetings were sometimes considered inefficient, especially when the team was inexperienced with Agile.
- Management aspects are also a concern. Many program managers in agile development are worried that upper-level management would ask for progress reports and productivity metrics that would be hard to gather in an agile work environment. Upper management still tries to get specific dates for specific deliverables.
- Another concern is coordinating with other teams. This is especially worrisome in larger projects where only a few groups are Agile and the rest are using a typical traditional model of software development. Problems arise in the scheduling of deliverables between dependent projects.

VII. Conclusions

This paper discusses about the Agile model with its characteristics, merits and demerits. It further throws a light on various approaches of Agile Methodology like Extreme Programming, Scrum, Feature Driven Development and Crystal Methods.

References

1. Kaushal Pathak and Anju Saha, "Review of Agile Software Development Methodologies," International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 2, February 2013
2. Gurleen Singh and Tamanna, "An Agile Methodology based Model for software development," International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 6, June 2014
3. J. Highsmith, Adaptive Software Development. New York, NY: Dorset House, 1999.
4. A. Cockburn, Agile Software Development. Reading, Massachusetts: Addison Wesley Longman, 2001
5. S. R. Palmer and J. M. Felsing, A Practical Guide to Feature-Driven Development. Upper Saddle River, NJ: Prentice Hall PTR, 2002.
6. K. Schwaber and M. Beedle, Agile Software Development with SCRUM. Upper Saddle River, NJ: Prentice-Hall, 2002.
7. Agile Processes in Software Engineering and Extreme Programming: 9Th International Conference, XP 2008, Limerick, Ireland, June 10-14, 2008 : Proceedings.
8. Mikio Aoyama, "Agile Software Process Model", IEEE Software, pp 454-455.

Evaluation of Black Hole Attack in Mobile Adhoc Network & Proposed Solution

Bhavna Sharma*

Abstract

Mobile ad hoc networks (MANETs) are extensively used in military and civilian applications. The dynamic topology of MANETs allows nodes to join and leave the network at any point of time. This generic characteristic of MANET has rendered it vulnerable to security attacks. In this paper, we address the problem of black holes attack & evaluate the simulation result on the parameters like Packet Delivery Ratio, Packet loss%, through put under the Black Hole Attack and proposed a method, based on establishing the additional route and key authentication, to avoid the black hole attack in mobile ad hoc network.

Key Words: MANET, Security attacks. Routing Protocols, AODV, Black hole NS-2, Cygwin

I. Introduction

Mobile ad hoc network is a collection of wireless mobile hosts forming a temporary network without the aid of any established infrastructure or centralized administration. In such an environment, it may be necessary for one mobile host to enlist the aid of other hosts in forwarding a packet to its destination, due to the limited range of each mobile host's wireless transmissions. In other words to make that work, typically each node need to act as a router to relay packets to nodes out of direct communication range. Under these circumstances, routing is much more complex than is conventional networks. Many of the possible solution are determined by the characteristics of the media, the behaviour of nodes and the data flow. Since research in ad hoc networking has resulted in such a large amount of routing algorithms and protocols, it has become more and more difficult to decide, which algorithm are superior to others under what conditions (E. Cheng, 2001).For the successful deployment. This is an important problem, since a wrong choice may have a severe impact on the performance, and consequently on the acceptance of new technology. Also providing just any protocol is not feasible, due to the different requirements on hardware and lower network layers. Further it would not make sense. Since all devices in an area would need to agree on one method of they want to communicate.

A mobile ad hoc network is a collection of digital terminals equipped with wireless transceivers that can communicate with on another without using any fixed networking infrastructure. Communication is maintained by the transmission of data packets over a common wireless channel. The absence of any fixed infrastructure. Communication is maintained by the transmission of data packets over a common wireless channel. The absence of any fixed infrastructure, such as an array of base stations, makes ad-hoc networks radically different from other wireless LANs. Whereas communication from terminal in an "infrastructure" network, such as cellular network, is always maintained with a fixed base station, a mobile terminal in an Adhoc network cans communication range. In other to transmit to a node that is located outside its radio range, data packets are relayed over a sequence of intermediate nodes using a

Bhavna Sharma*
Research Scholar, Poornima University

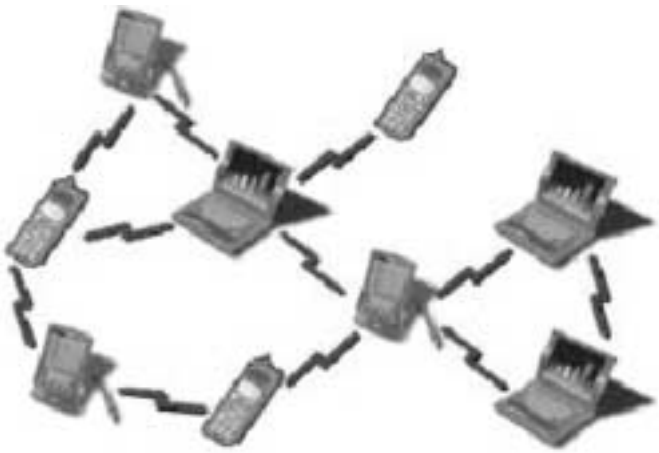


Figure 1: Infrastructure Network

store and forward "multihop" transmission principle. All nodes in an ad hoc network are requires to relay packets on behalf of a multihop wireless network.

II. Routing in MANET

Routing in ad-hoc network involves determining a path from the source to the destination data can be communicated and the delivery of the packets to the destination nodes while nodes in the network are moving freely. Due to this node mobility, a path established by a source may not exist after a short interval of time. To cope with node mobility, nodes need to maintain routes in the network. Routing protocols for ad-hoc networks broadly fall into pro-active, reactive, hybrid and location-based categories depending upon how nodes can establish and maintain paths.

Routing schemes can be classified into three categories namely, table driven (or proactive) routing protocols; On Demand (or Reactive) Routing protocols and hybrid (Location Based) routing protocols.

(a) Table Driven Routing (Global/Proactive)
Examples of ProactiveRouting Protocols are:
Global State Routing (GSR).
Hierarchical State Routing (HSR).
Destination Sequenced Distance Vector Routing (DSDV).

(b) On Demand Routing (Reactive Protocols)
Examples of reactive protocols are:

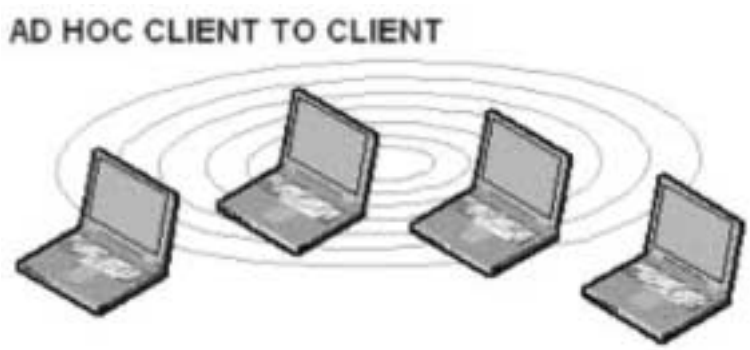


Figure 2: Infrastructure less Network

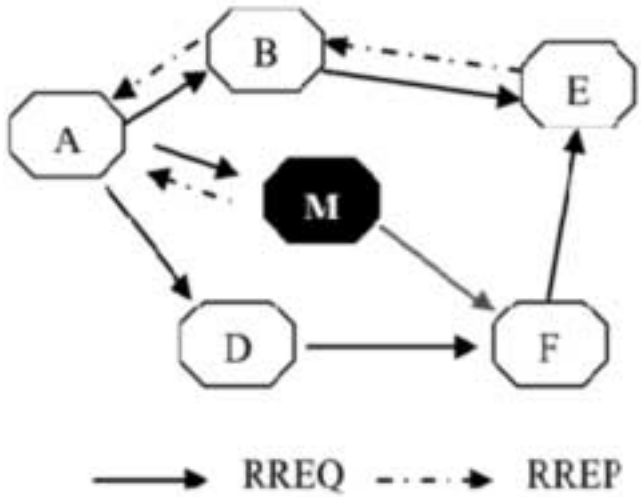


Figure 2: Black hole Attack in MANET

Ad hocOn-demand Distance Vector Routing (AODV).

Dynamic Source Routing (DSR).

Location Aided Routing (LAR).

Temporally Ordered Routing Algorithm(TORA).

(c) Hybrid Routing

There exist a number of routing protocols of globally reactive and locally proactive states. Hybrid routing algorithms ideal for Zone Based Routing Protocol (ZRP).

III. Black Hole Attack on MANET

Routing protocols are exposed to a variety of attacks. Black hole attack [7] is one such attack and a kind of Denial Of Service (DoS) [8][9] in which a malicious node makes use of the vulnerabilities of the route discovery packets of the routing protocol to advertise itself as having the shortest path to the node whose packets it wants to intercept [10][11]. This attack aims at modifying the routing protocol so that traffic flows through a specific node controlled by the attacker. During the Route Discovery process, the source node sends RREQ packets to the intermediate nodes to find fresh path to the intended destination. Malicious nodes respond immediately to the source node as these nodes do not refer the routing table. The source node assumes that the route discovery process is complete, ignores other RREP messages from other nodes and selects the path through the malicious node to route the data packets. The malicious node does this by assigning a high sequence number to the reply packet. The attacker now drops the received messages instead of relaying them as the protocol requires.

In the above figure 3, imagine a malicious node „M . When node „A broadcasts a RREQ packet, nodes „B „D and M receive it. Node M , being a malicious node, does not check up with its routing table for the requested route to node „E. Hence, it immediately sends back a RREP packet, claiming a route to the destination. Node A receives the RREP from „M ahead of the RREP from „B and „D. Node „A assumes that the route through „M is the shortest route and sends any packet to the destination through it. When the node „A sends data to „M, it absorbs all the data and thus behaves like a „Black hole .

In AODV, the sequence number is used to determine the freshness of routing information contained in the message from the originating

node. When generating RREP message, a destination node compares its current sequence number, and the sequence number in the RREQ packet plus one, and then selects the larger one as RREPs sequence number. Upon receiving a number of RREP, the source node selects the one with greatest sequence number in order to construct a route. But, in the presence of black hole when a source node broadcasts the RREQ message for any destination, the black hole node immediately responds with an RREP message that includes the highest sequence number and this message is perceived as if it is coming from the destination or from a node which has a fresh enough route to the destination. The source assumes that the destination is behind the black hole and discards the other RREP packets coming from the other nodes. The source then starts to send out its packets to the black hole trusting that these packets will reach the destination. Thus the black hole will attract all the packets from the source and instead of forwarding those packets to the destination it will simply discard those. Thus the packets attracted by the black hole node will not reach the destination.

IV. Proposed Work

We propose an additional route to the intermediate node that replies the RREQ message to check whether the route from the intermediate node to the destination node exists or not. When the source node receives the FurtherReply (FRp) from the next hop, it extracts the check result from the reply packets. If the result is yes, we establish a route to the destination and begin to send out data packets. If the next hop has no route to the inquired intermediate node, but has a route to the destination node, we discard the reply packets from the inquired intermediate node, and use the new route through the next hop to the destination. At the same time, send out the alarm message to whole network to isolate the malicious node. If the next hop has no route to the requested intermediate node, and it also has no route to the destination node, the source node initiates another routing discovery process, and also sends out an alarm message to isolate the malicious node. Thus we avoid the black hole problem, and also prevent the network from further malicious behaviour. But here we assume the black hole nodes do not work as a group and propose a solution to identify a single black hole. However, the proposed method cannot be applied to identifying a cooperative black hole attack involving multiple nodes. We may also develop a methodology to identify multiple black hole nodes cooperating as a group. The technique works with slightly modified AODV protocol and makes use of the Data Routing Information (DRI) table in addition to the cached and current routing tables. A black hole has two properties. First, the node exploits the ad hoc routing protocol, such as AODV, to advertise itself as having a valid route to a destination node, even though the route is spurious, with the intention of intercepting packets. Second, the node consumes the intercepted packets.

The solution that we propose here, basically, only modifies the working of the source node without altering intermediate and destination nodes. In this method two main things are added namely Data Routing Information table and cross checking.

Proposed Algorithmic Approach
Steps:
1: Source node broadcasts RREQ
2: Source node receives RREP
3: if RREP is from destination or a reliable node Then route data packets (source route)
4: Else
{

Send further request and identity of intermediate Node to next Hop node

Receive further request, next Hop node of current next hop node, Data Routing

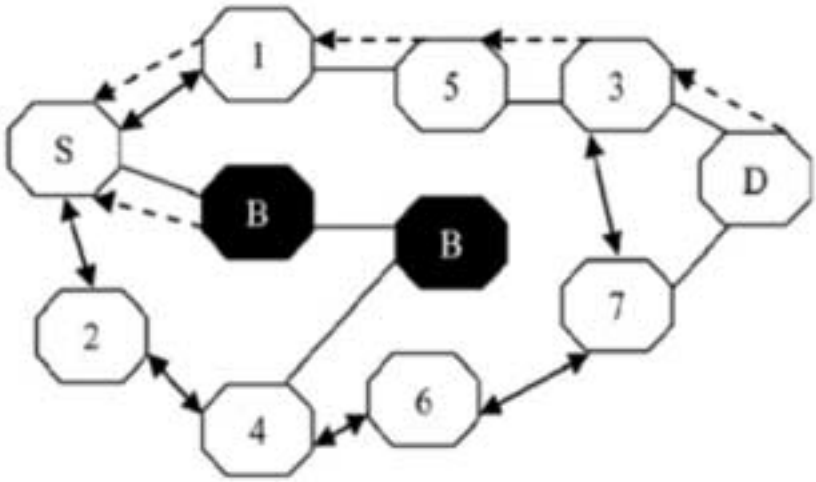


Figure 4: Solution to Avoid Black Hole Attack

Information entry for next hope nodes next hop. Put a data routing information entry for current intermediate node.

5: if (next hop node is a reliable node)

{

Check intermediate node for black hole using data routing information entry

if (intermediate node is not a black hole)

route data packets (source route)

else

{

Insecure route

Intermediate node is a black hole

All the nodes along the reverse path from intermediate node to the node that generated RREP are black holes (i.e. a malicious node)

}

}

else

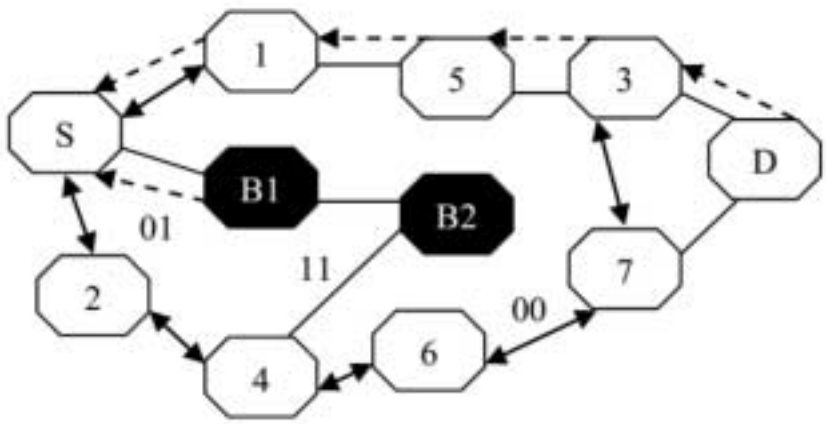
Current intermediate node = next hop node

}

6: Repeat step 4 & 5 until intermediate node is not a reliable node

Working Principle

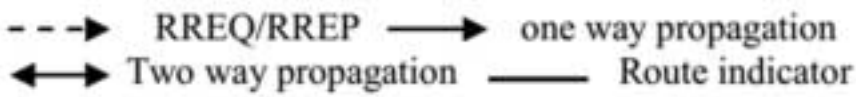
The solution to identify multiple black hole nodes acting in cooperation [12] involves two



bits of additional information from the nodes responding to the RREQ of source node S. Each node maintains an additional Data Routing Information (DRI) table

In the DRI table, 1 stands for „true and 0 for „false. The first bit "From" stands for information on routing data packet from the node (in the Node field) while the second bit "Through" stands for information on routing data packet through the node (in the Node field). In reference to the example of Figure 3, a sample of the database maintained by node 4 is shown in Table 1. The entry 1 0 for node 3 implies that node 4 has routed data packets from 3, but has not routed any data packets through 3 (before node 3 moved away from 4). The entry 1 1 for node 6 implies that, node 4 has successfully routed data packets from and through node 6. The entry 0 0 for node B2 implies that, node 4 has NOT routed any data packets from or through B2.

We define the following convention for protocol representation in fig 4 & 5.



Cross Checking: In our techniques we rely on reliable nodes (nodes through which the source node has routed data) to transfer data packets. The modified AODV protocol and the algorithm for our proposed methodology are illustrated in Figure 5. In the protocol, the source node (SN) broadcasts a RREQ message to discover a secure route to the destination node. The Intermediate Node (IN) generating the RREP has to provide its Next Hop Node (NHN) and its DRI entry for the NHN. Upon receiving RREP message from IN, the source node will check its own DRI table to see whether IN is a reliable node. If source node has used IN before to route data, then IN is a reliable node and source node starts routing data through IN. Otherwise, IN is unreliable and the source node sends FRq message to NHN to check the identity of the IN, and asks NHN: 1) if IN has routed data packets through NHN, 2) who is the current NHN s next hop to destination, and 3) has the current NHN routed data through its own next hop. The NHN in turn responds with FRp message including 1) DRI entry for IN, 2) the next hop node of current NHN, and 3) the DRI entry for the current NHN s next hop. Based on the FRp message from NHN, source node checks whether NHN is a reliable node or not. If source node has routed data through NHN before, NHN is reliable; otherwise, unreliable. If NHN is reliable, source node will check whether IN is a black hole or not. If the second bit (i.e. IN has routed data through NHN) of the DRI entry from the IN is equal to 1, and the first bit (i.e. NHN has routed data from IN) of the DRI entry from the NHN is equal to 0, IN is a black hole. If IN is not a black-hole and NHN is a reliable node, the route is secure, and source node will update its DRI entry for IN with 01, and starts routing data via IN. If IN is a black -hole, the source node identifies all the nodes along the reverse path from IN to the node that generated the RREP as black hole nodes. Source node ignores any other RREP from the black holes and broadcasts the list of cooperative black holes. If NHN is an unreliable node, source node treats current NHN as IN and sends FRq to the updated IN s next hop node an d goes on in a loop from steps 4 through 6 in the algorithm.

V. Simulation Environment

We have implemented Black hole attack in an ns-2 [13] simulator. For our simulations, we use CBR (Constant Bit

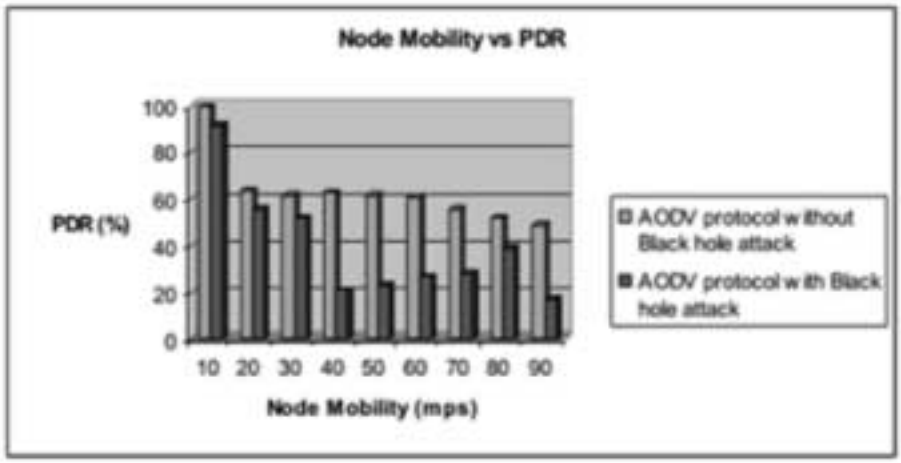


Figure 6: Solution to Avoid Black Hole Attack

Rate) application, TPC/IP (full duplex communication), IEEE 802.11b MAC and physical channel based on statistical propagation model. The simulated network consists of 30 randomly allocated wireless nodes in a 500 by 500 square meter flat space. The node transmission range is 250-meter power range. Random waypoint model is used for scenarios with node mobility. The selected pause time is 30s seconds. A traffic generator was developed to simulate constant bit rate (CBR) sources. The size of data payload is 512 bytes. In our scenario we take 30 nodes in which nodes 1-22 and 25-30 are simple nodes, and node 23 and 24 are malicious node or Black hole node. The simulation is done using ns-2, to analyze the performance of the network by varying the nodes mobility [14] [15]. The metrics used to evaluate the performance are given below.

a) Packet Delivery Ratio: The ratio between the number of packets originated by the "application layer" CBR sources and the number of packets received by the CBR sink at the final destination.

(b) Throughput: Throughput is the average rate of successful message delivery over a communication channel.

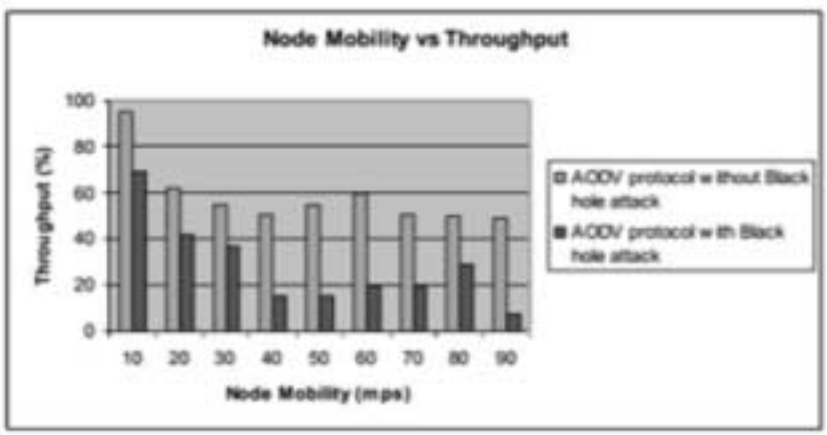


Figure 7: Solution to Avoid Black Hole Attack

(c) Node Mobility: Node mobility indicates the mobility speed of nodes.

VI. Result & Discussion

The fig.6 shows the effect to the packet delivery ratio (PDR) measured for the AODV protocol when the node mobility is increased. The result shows both the cases, with the black hole attack and without the black hole attack. It is measured that the packet delivery ratio dramatically decreases when there is a malicious node in the network. For example, the packet delivery ratio is 100% when there is no effect of Black hole attack and when the node is moving at the speed 10 m/s. but due to effect of the Black hole attack the packet delivery ratio decreases to 82 %, because some of the packets are dropped by the black hole node.

It is observed from the fig.7 that, the impact of the Black hole attack to the Networks throughput. The throughput of the network also decreases due to black hole effect as compared to without the effect of black hole attack. We vary the speed of the node and take the result to the different node speed.

VII. Conclusion and Future Work

In this paper we have gone through the routing security issues of MANETs, described the black hole attack that can be mounted against a MANET and proposed a feasible solution for it in the AODV protocol. The proposed solution can be applied to a) Identify single and multiple black hole nodes cooperating with each other in a MANET; and b) Discover secure paths from source to destination by avoiding multiple black hole nodes acting in cooperation. Also we showed that the effect of packet delivery ratio and Throughput has been detected with respect to the variable node mobility. There is reduction in Packet Delivery Ratio and Throughput. In Black hole attack all network traffics are redirected to a specific node or from the malicious node causing serious damage to networks and nodes as shown in the result of the simulation. The detection of Black holes in ad hoc networks is still considered to be a challenging task. We simulated the Black Hole Attack in the Ad-hoc Networks and investigated its affects. In our study, we used the AODV routing protocol. But the other routing protocols could be simulated as well. All routing protocols are expected to present different results. Therefore, the best routing protocol for minimizing the Black Hole Attack may be determined

References

1. W. Stallings, Wireless Communication and Networks, Pearson Education,2002.
2. C. Perkins, Ad Hoc Networks, Addison-Wesley, 2001.
3. Y. Hu, D. Johnson, and A. Perrig, SEAD: Secure Efficient Distance Vector Routing in Mobile wireless Ad-Hoc Networks. Proc. of the 4th IEEE workshopon Mobile Computing Systems and applications (WMCSA'02).
4. A. Tanenbaum, Computer Networks, PH PTR, 2003.
5. Abhay Kumar Rai, Rajiv Ranjan Tewari & Saurabh Kant Upadhyay, "Different Types of Attacks on Integrated MANET-Internet Communication," International Journal of Computer Science and Security (IJCSS) Volume: 4 Issue: 3.
6. Ping Yi, Yoe and Futai Zoo and Ning Liu, " A survey on Security in Wireless Mesh Networks", Proceedings of IETE Technical Review, Vol.27 Jan-Feb 2010
7. Lidong Zhou J. Haas, "Securing Ad-hoc Networks", IEEE Network Magzine, 13,6, pages 24-30, 1999.

Comparative Analysis of Digital Payment Modes in India

Shubham Kapoor*
Sakshi Singh**
Simranjeet Kaur***

Abstract

The only mode of doing a transaction in India is cash i.e. less than 5% of all the payments are done electronically. But in 2016 the budget speech talked about the idea of making India a cashless nation with the vision of block the flow of black money. To fulfill the aim of eradicating the black money even RBI is also taking certain steps and providing facilities to make the digital payments more accessible and more secure. An unveiled document by RBI "Payment and settlement system in India: Vision 2018" reveals the plan made by RBI to encourage payments and enables India to move towards the cashless society. In this research papers we are going to discuss the modes which are provided and developed by government or private companies to encourage more cashless transactions and make them more secure.

Key Words: Embedded System, Smart Card Technology (SCT), Encrypted Storage, UPI, USSD, AEPS.

I. Introduction

After the introduction of money as medium of exchange of goods and services fraud and volume payment became a serious issue. Before that, in the days of barter system of in other words "double coincident of wants" one is always worried about of finding the correct person who wants the same service he wants to exchange for. When gold, silver and bronze pieces began to be produced, it became unsafe to keep gold at home and it was absolutely not safe to travel with a wallet full of valuables. Robbery was the serious threat those days and it was also difficult to carry the gold every time. Coming to today world using only cash as a medium of exchange is still a serious issue i.e. it is not secure and also there is more problems such as flow of black money or making and using of fake currency. These all problem creates a chaos in markets and make the transactions unsuccessful. To overcome all these issues the idea of cashless transaction was introduced which make an individual to pay more amount of money without the use of cash and with more security. This all can be done with the use of banks and making banks connected 24X7 to the consumer. As there are benefits of it still there lies some security issues such as hacking of systems in banks, reviling of confidential data, data breaching etc.

Shubham Kapoor*
Student BCA
IINTM

Sakshi Singh**
Student BCA
IINTM

Simranjeet Kaur***
Assistant Professor
IINTM

II. What is Cashless Transaction

Cashless transactions means transferring of money from one person account to another person account with the use of digital cards or any other digital medium. Cash less payments can be done when the banks are connected to the person with the certain medium which can be mobile phones. Many of the certain companies are working of making the cashless transactions more secure and to make them accessible

to more and more people. Many of the mobile applications are being made and made which can be connected to the bank of the consumer and with the permission of the consumer the funds can be transferred from one account to another account. The following are the benefits of cashless transactions:

- If cashless transactions are widely accepted it would be convenient for everyone as there will be no hassle and it will be less risky to carry cash in more amounts.
- No need to memorise the payments or expenses as all the transactions are recorded automatically.
- Tax rates can be reduced.
- Cashless transaction are easily traceable therefore it blocks the use of black money and fake currency which are the major threat to market and encourage people to pay taxes regularly.

Along with some advantages there lie some drawbacks also as follows.

- It is not safe because banks servers can be hacked within few hours and confidential data of consumer can be taken off.
- Data breaching and data theft are the common problems.
- Not so simple to use.

III. Security Issues in Digital Payments

The weakest security link in any transaction is not the technology it is the user and the lack of understanding the security issues. To get to know this some people give others their card and PIN to withdraw money from ATM. The main difference between the cash and digital is that in the cash there is the limit to the loss of the notes or the number of notes in digital the risk is more as in wallets there is the advantage that one could transfer money bit by bit and lower the risk of exposure. That of course doesn't mean that digital shouldn't be an option. Also there is no privacy in the cash less which means that all the transactions are being documented and also the transaction are traced. As it gives the advantage to the government for the taxation but there is no protection for the citizen who owns the data.

IV. Modes of Cashless Payments

Some of the modes for making payments without the use of cash are as follows.

A. Banks cards

A bank card is a small plastic card which is issued by a bank to its costumer through which a costumer can make payments or perform any kind of transactions. It contains a magnetic strip which stores the confidential data of the costumer and by swiping it on the card reader machine the machine reads the codes on the magnetic strips and make the link between the bank and the place where the funds are to be transferred. The card also contains the costumer name and a 16 digit unique card number validity date nod expiry date along with the unique CVV number.

- Types of banks cards.

There are certain types of cards which are provided by banks stated below.

Debit card
Credit card
ATM card
gift card

B. Unstructured supplementary service data (USSD)

Unstructured supplementary service data (USSD) which is also sometimes referred as "Quick Codes" or "Featured Codes" is the protocol which is used by the GSM cellular network users to communicate with their service provider. Nowadays, this service is also provided by the banks to their customers to connect to the bank. In this process a customer needs to register a mobile number while opening of account. USSD is used by banks mainly to enable mobile banking. As we are moving towards cashless society we need to perform every transactions in cashless mode and Mobile banking is one of the mode of cashless transaction. Mobile banking can be a saviour when we don't have internet access. The only requirements for activations are a bank account and any mobile on cellular network.

According to the RBI, the USSD payments method can be used for sending money up to the limit of Rs.5,000 per transaction. Per transaction will charge

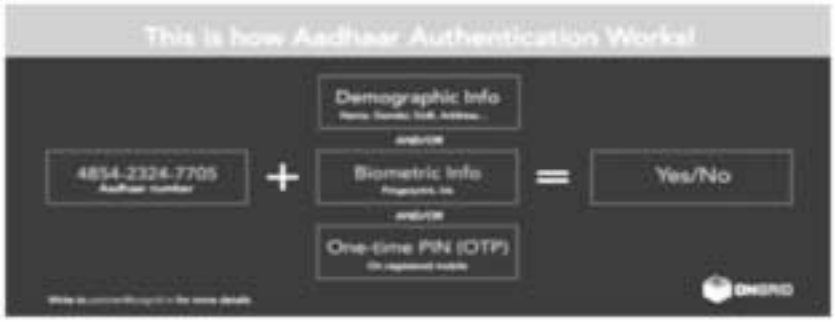


Fig. 1: working of aadhar enabled payment service

Source: https://upipayments.co.in/aadhaar-enabled-payment-system/

Rs. 0.50 which will be added in mobile bill. This service is available 27*7 and 365 days. For getting started with the USSD consumer need a phone, a bank account and the mobile number registered for mobile banking. Once all the registrations are done a Mobile Money Identifier (MMID) will be generated. It is a random 7 digit number issued by the banks. Along with the MMID a MPIN which is a four digit code to authorize the transaction will be generated. And once you get the MPIN consumer can change it to any digits of choice.

http://www.bgr.in/news/how-to-use-ussd-based-mobile-banking-heres-everything-you-should-know/

C. Aadhar Enabled payment system

Another way of getting money from the bank is AEPS. In this system of getting money from the bank doesn't require any signature or the debit card. A consumer doesn't even need to go to the bank for getting money by using aadhar enabled payment system. Rather it uses aadhar data for the authentication. It is the initiative by the NPCI. In order to increase the online transaction in the country AEPS is introduced which includes the working groups on MicroATM's standards & central infrastructure. NPCI proposed to conduct a lab level proof of concept (poc), integrating the authentication and encryption standards of UIDAI to ensure MicroATM and transactions using aadhar before they are put in the actual use.

Some of the benefits of using AEPS are as follows.

Consumer can easily perform the financial and non financial transactions using it and it can also be used by the people in rural areas also.

It is fast and secure as the biometric is used i.e. is the finger print it can never be copied and can never be forgotten.

No need of any kind of signature or debit card.

- Transactions Through AEPS

The AEPS gives the banking facilities which are only basic such as.

Balance check

Aadhar to aadhar funds transfer

Cash withdrawal

Cash deposit

- Requirements for AEPS

Before making any kind of transactions using the AEPS it must be registered with the bank account. If it is not linked then it can't be used

- Why Aadhar For Authentication

The biometric data such as the finger print scan and the iris image gets linked to the aadhar number of the person, as the biometric is unique and no two person can have same fingerprint they can't be duplicated and can't be used in some kind of frauds. Thus aadhar can be verified by the finger prints. Aadhar can be used as the perfect proof for the banking transactions.

- Working Of Aadhar Authentication

For using the aadhar number for any kind of transaction first the bank official will take the finger prints in the machine. The machine then scan the finger print and give whole details of the person whose finger print is being scanned on the aadhar payment app. The app immediately tallies the fingerprints for

Fig.2: UPI

Source:

the transactions and if there is any kind of mismatch the system will deny the for the transactions and the transaction will fail. The more broad based diagram of the whole process is shown in fig 1.

- Charges Of AEPS

AEPS can cost up to Rs.15 per transaction and it is costly than UPI as it free of cost. For AEPS there are 3 institutions involved which charges the fees such as/

Nominal fees for authentication can be charged by UIADI

NPCI charges 10 paisa for authentication and 25 paisa for settlement.

1% of the transaction is charged by the bank and the minimum fees for the other banks would be Rs 5 and maximum fees would be Rs. 15.

But as AEPS may be costly than UPI and USSD but it gives a lot of convenience to the rural people. AEPS helps to provide banks to the door steps of the people and helps in saving a lot of time and transport expense. It is similar to the ATM at every doorstep hence the government calls the POS as MicroATM's.

D. Unified Payment Interface (UPI)

UPI is a system which is empowers multiple banking services features like funds transfer and merchant payment just by using a single application in the mobile phones. It was launched by the national payment corporation of India to promote cashless transaction in India. UPI has again made the payment system more convenient for the people as the people can now just easily put the bank in their pockets and can easily perform all the task which they does in bank on their mobile phones only. Also the UPI is unique as it immediately transfers money through mobile device round the clock 24*7 and 365 days. And also many other features can be done like bill sharing with friends, donations, collections, barcode based payments. Also there are many benefits by using the UPI to the both banks and the costumer such as

- Benefits To The Banks

UPI helps to enables the seamless merchant transactions.

This method is safer and secured and as it is developing more and more features can be added hence it is innovating.

There is a universal application for the transaction.

On just the single click there are the 2 factor authentications.

- Benefits To The Consumer

It is available all the time i.e. 24*7 and 365 days and any user can use it any time anywhere the only requirement is the stable internet connection.

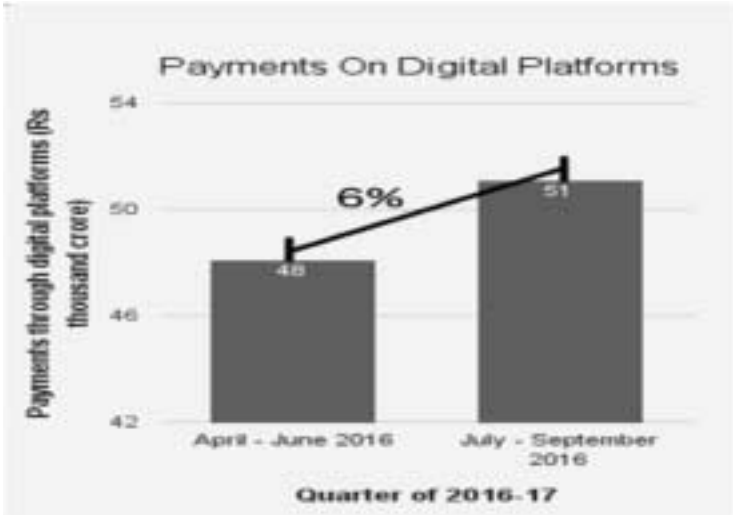


Fig 3: payment through digital platforms.

Source: https://upipayments.co.in/aadhaar-enabled-payment-system/

There is the single click authentication

By using a single application different banks account can be accessed.

- Sending Money On UPI

First of all for sending money the user choose the Send money option as shown in the fig.2 and then the payee's virtual payment addressed is being entered. After it the amount is entered and then the remark for the transaction are been written which means for which the transaction is been made or any kind of record for the transaction and then the details are confirmed after the confirmation the person hit the send button and then the money is send. Whole process of sending the money is being shown in fig2.

- Advantages Of UPI

UPI gives us many advantages as given follows

It is the cheapest way of transferring of money as NEFT charges minimum of RS. 2.5 For a transfer but the UPI charges less than 50 paisa for a transaction. And just because of this low cost the UPI has promoted the non cash or cashless transaction of small amounts also.

As it is a cashless transaction it makes the user's free form the cash and also UPI makes the small transaction more feasible and smooth.

It is fast as in the online transactions and the card payment the system ask for the account details and CVV number and the validity period it is very time consuming and also bit insecure and this problem is solved by the UPI the payment through UPI doesn't require the card details the costumer only gives the virtual payment address and it is similar to the email ID.

It is more secure than the present mode of transaction as in the payment the person never shares the bank details such as CVV or card number and the plus point is that authentication is done on the smart phones only and it is done by punching the 2 different pins.

UPI is based on IMPS platform hence it works 24*7 and the payment can be made any time and there problem on holidays as well.

E. E-Wallets

E-Wallets are the encrypted storage medium which holds the credit card and other confidential payment information that can be used for the transaction or any type of fund transfer. It is the system that stored the consumer data which can be easily retrieved at the time of payment. Also e-wallets offers the powerful protection for the data and the programme uses the strongest encryption which is used by the US government permits which means using the RC4 encryption algorithm with a 56 bit key.

Fig 4: private and government expenditure on digital payments.
Source: https://upipayments.co.in/aadhaar-enabled-payment-system/

- How Does E-Wallets Works

E-wallets are the small browser extensions that users carry in their mobile form or can access them on the websites also and if the consumer wants to make any kind of payment it first browse the site and after making the selection user just select the option of E-wallet in the task bar and then the wallet automatically fills the necessary data and the site then transaction once the data entered gets verified. Typically the digital wallets are stored on the client side and are easily self-maintained and fully compatible with most of the e-commerce websites. Now a day's server side e- wallets are gaining more and more popularity among major retailers due to some security, efficiency, and also provides end-user which increases the satisfaction of their overall purchase. The M-PESA mobile payments system are the micro financing services have widespread use in Kenya.

- Advantages Of E-Wallets

As many of the online shopping payments gets cancelled due to frustration in filling the forms the e-wallets combats this problem by giving the users the option to transfer their confidential transaction information securely and accurately. This has made the transaction more simple and fast.

E-wallets are secure and the transaction is made securely and doesn't pass the payment details to the websites. Increasingly this approach is a feature of online payment gateways.

For those who are having the smart phones can easily have their w-wallets from Google play store or appstore for iphone also these apps work at 2g speed also hence they are available 24*7 and everywhere.

- Shifting From Cash To Cashless

As after the demonetization drive in November there were over 100 million users and 1 lack retail touch points across 80 thousand villages and 630 cities till the first week of November were using online transaction after that we have seen 25% jump in new users which are mostly from rural and semi urban areas. Still over 93% of people in rural areas have not done any type of online transaction. Shifting from cash to basic cash less can be seen in the two graphs shown below.

Keeping all the positive and negative aspects in mind of cashless everyone is now shifting from cash to cashless which is also helping the government for collecting taxes and eradicating the black money also there are further more developments going on to resolve the security issues.

Making the online and cashless payments accessible to everyone and to make them free of cost various applications such as BHIM app is also is trend now and will be in more broader use till the end on 2017.

Table 1: Functionality Analysis

S. No.	Functio-nality	AEPS	USSD	UPI	E-Wallets	Bank Cards
1.	Ease of use	Easy to use as it only requires the finger print.	Not easy to use as the client has to remember the complex USSD codes	Easer to user as it is just a mobile application to be installed	Easy to use	Easy to use but user have to go to an ATM to have cash.
2.	Speed	Fast	Fast	Fastest	Faster than any other	Faster payments but dependent on server of bank.
3.	Security	Secured with biometric that is the iris scan and the finger print.	Secured by the pin code.	Secured	user has to enter the pins twice to open an app or to make any kind of payment	Less secure as the pin can easily been seen
4.	Availability	Not available 24*7	Available 24*7	24*7 available	24*7 available	Available 24*7
5.	Knowledge/ language barrios	No knowledge barrio as it is widely used in rural areas.	No knowledge barrior.	There is a language barriors as the person needs a Smart-phone and has to be educated	Only in English	No language barriors
6.	Cost/ expense	Cost up to Rs. 15 per transaction	Cost 50 paisa per transaction	No cost	No cost cutting	No cost.
7.	User preference	Proffered in rural are than the urban	Proffered by people	User prefers as it is fast secured and easy	User prefers e-wallets for payments as it is faster and also it provides them certain offers on shopping	Many of the users now days prefer credit cards as it also provides them certain offers
8.	Payment Scale	Can pay up to Rs. 5000.	Can transfer up to Rs. 5000	Can pay huge amounts	Can pay up to 5000 per transaction.	From ATM one can withdraw up to 35000 and up to 50000 payments can be done.

V. Functionality Analysis

Keeping all the positive aspects of digital payments let's see the comparison on the basic terms in which they can be differentiated such as ease of use which means how easy it is to use that type of payment mode, speed that means how fast the transaction is, how secure the mode is, there is a language barrier not, cost cutting (if any), then he user preference and then the payment scale that is how much amount can be paid by the particular mode of payment. All of the above are mentioned in the table no.1

VI. Conclusion

There are many benefits of cashless transactions but it has many loop holes also such as security issues. As India is moving towards a cash less society hence citizens of India must be aware of all the scams and all the treats related to the online transaction and if it's not going to take place then the cash will always be the king and the problem of black money and fake currency can never be solved Currently, AEPS system seems to be the leading system for the transactions and to make it more secure the government has to make major developments.

References

1. https://upipayments.co.in/aadhaar-enabled-payment-system/
2. http://www.npci.org.in/aepsoverview.aspx
3. https://upipayments.co.in/aadhaar-enabled-payment-system/

National Conference on Innovative Research in Computer Science and Information Technology

The purpose of the Conference is to bring together the academicians, researchers, scientists, engineers and practitioners from various institutions, universities, colleges and industry to discuss the relevance of various futuristic technologies emerging in Networking, Big Data, Data Mining, Security and also areas related to innovations in Information Technology throughout the world. This Conference is aimed to expand its coverage in the areas specified where expert talks, young researcher's presentations will be placed in every session will be inspired and keep up your enthusiasm.

Conference Objectives

- To present the latest research and results of experimental work related to Computer Science and Information Technology
- To endow opportunities for delegates to exchange new ideas and application experiences.
- To address the real-world opportunities and challenges faced by today's IT professionals.
- To promote high level of interaction among stakeholders in theoretical, experimental and applied concepts.

Conference Themes

Topics covered by the Conference include, but are not restricted to, the following

Communication and Networks Big Data and Data Mining

- Mobile Ad hoc and Wireless Sensor Network
- Communication Network Architecture
- Communication Network Protocols
- Next Generation Network Technologies
- Future Internet Architecture and Protocols
- Ubiquitous Networks and Cloud Computing
- Wireless and Mobile Computing
- Tracking and Locating Mobile Users
- Broadband Access Networks
- Energy Saving Protocols
- Data and Information Quality for Big Data
- Big Data Infrastructure
- Cloud/Grid/Stream Computing for Big Data
- Big Data Search and Mining
- Social Web Search and Mining
- Data Mining Tools and Software's
- OLAP Technologies
- Big Data Tools
- Data Warehouse Management
- Data Visualization

Data Security

- Analysis of security protocols
- Authentication and authorization of Systems
- Botnets
- Information Assurance and Security
- Forensics and Diagnostics for Security
- Intrusion Detection and Prevention
- Cyber Security
- Web Security, Client and Server Security
- Applications of Cryptographic Techniques
- Self-protecting and Self-healing System

Multifaceted Areas

- Computer Architecture and Engineering
- Artificial Intelligence and Robotics
- Soft Computing
- Software Engineering
- Human Computer Interaction
- Internet of Things (IoT)
- Social Networking
- Crowd Sourcing
- Image Processing
- Theoretical Computer Science

Institute of Innovation in Technology and Management

Affiliated to GGSIP University, 'A' Grade by GNCTD & Assessed as 'A+' by SFRC
NAAC Accredited Institute
(ISO 9001-2008 Certified Institute)

D-27/28, Institutional Area, Janakpuri, New De1hi-110058
Tel: 011-28520890, 28520894, Te1e Fax: 011-8520239
E-mail: director@iitmjp.ac.in Website: http://www.iitmjp.ac.in