Improved greedy routing protocol for VANET

Ambica

Department of computer science

BGIET ,Sangrur

Punjab,India

Avinash Jethi

Department of computer science

BGIET ,Sangrur

Punjab,India

Abstract :VANET (vehicular ad-hoc network) is a classification of MANET in which vehicles act as mobile node and provides a different approach to Intelligent transport System (ITS). VANET is an emerging area for Intelligent Transportation System (ITS) which can result in increased traffic safety, collision warning through exchange of messages through wireless media. Efficient routing protocols are required for efficient communication among vehicles. In the given paper, we surveyed various VANETs protocols and along with its advantages and disadvantages. We compare our proposed protocols via NS-2 based simulations and show the performance of different protocols.

keywords: VANET ,routing protocols, GPRS,GSPR-MV,V2V

1. INTRODUCTION

Vehicular ad hoc network (VANET) is becoming a promising area in the field of wireless network. VANET is a subset of MANET in which vehicles act as mobile node that provides communications among nearby vehicles, and between vehicles and nearby road side equipments through wireless medium. The U.S. Federal Communications Commission (FCC) has recently allocated the 5.85- 5.925 GHz portion of the spectrum to inter-vehicle communication (IVC) and vehicle-to-roadside communication (VRC) under dedicated short-range communications (DSRC). The vehicles and the associated infrastructure (all nodes) are equipped with wireless devices to gather data, and the data is processed to determine present traffic conditions and disseminate it over longer distances and can use it to alter its decision. For example they can change their route based on the information received in case of any congestion or collision warning. Regular enhancement in infrastructure of VANETs can result in increased safety and comfort of passenger and driver. VANET approach is scalable and has low maintenance costs. Government is taking much interesting in VANET because of safety it provides. Many projects such as COMCAR [15] , CarTALK [16], CarNet [17]. Characteristics of VANETs are [14] -

- Frequent Disconnected Network

- Dynamic topology.

- Battery Power and Storage Capacity

- Communication Environment

- Interaction with onboard sensors

VANET provide communication with vehicles in three modes
- Inter-vehicle communication
- Vehicle-to-roadside communication, and
- Routing-based communication
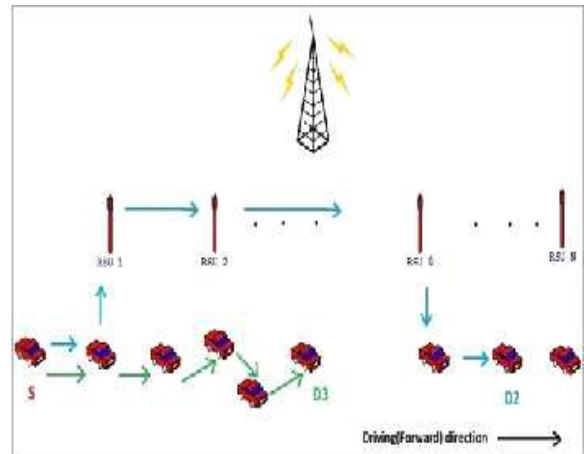
VANET routing protocols are usually designed for urban environments where vehicles are equipped with wireless connectivity and a GPS device. Topology is dynamic and the routing protocol must be fast enough to find a route in order to have a good response time. Based on the location of vehicle list of its neighbors is generated which are then used for discovery of route from source to destination. The consumption of battery in vehicles is very fast so in order to improve performance and energy conservation in fast moving vehicles light weight routing

protocols need to be defined in Physical and Data link layer. For discovering the route ad hoc network use two different types of protocols which are classified as **Proactive** and **Reactive [**18] protocols. **Proactive** protocols maintain the routing information from each node to every node at all the time and that is up-to-date. **Reactive** protocols, also called "on demand" node initiates a route discovery process only when a route to destination is required.

In this paper, Section II describes the VANET architecture. Section III presents a survey of these protocols experimented on to VANET and their advantages and disadvantages.

2. VANET NETWORK ARCHITECTURE

Architecture of vanet network can be classified into 3 categories: cellular WLAN, pure ad hoc and hybrid.

2.1 Cellular /WLAN

Fixed cellular gateways and WLAN/WiMAX access points are used at traffic intersection to connect to internet and gather information. VANET combines both cellular and WLAN network so that WLAN is used where an internet connection is available. But it adds to cost because of fixed gateways and other hardware devices. Figure 1 shows cellular/WLAN architecture.
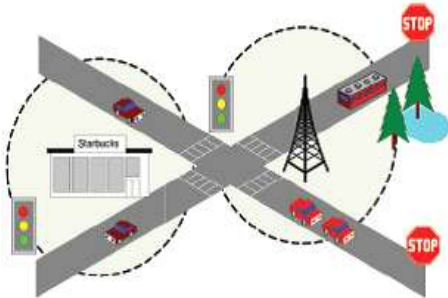

Figure 1 Cellular/WLAN Network Architecture

2.2 Ad hoc

To overcome the shortcoming of cellular/WLAN network vehicles can form an ad hoc network among themselves. Figure 2 shows an ad hoc network. It helps in vehicle to vehicle communications and achieves certain goals, such as blind crossing.
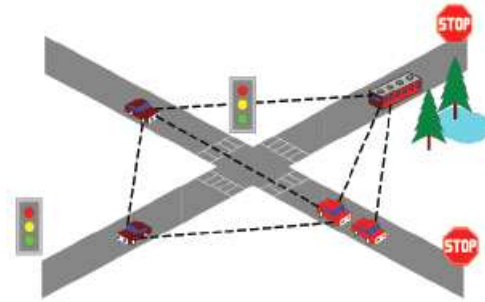

Figure 2 Ad hoc Network

2.3 Hybrid

Hybrid architecture is a combination of infrastructure and ad hoc network. It uses vehicles with both WLAN and cellular capabilities as gateways and mobile network routers so that vehicles with only WLAN capability can communicate with them through multi-hop links to remain connected to the world. The hybrid architecture can provide better coverage, but also causes new problems, such as the seamless transition of the communication among different wireless systems.
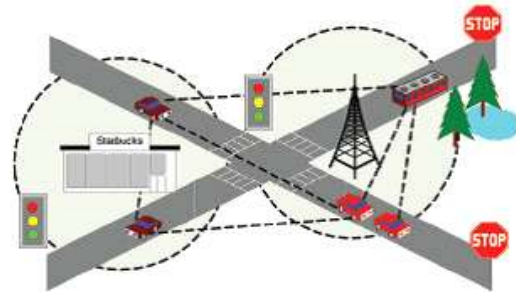

Figure 3 Hybrid Architecture

3. ROUTING PROTOCOLS IN VANET

Routing protocols for VANET are designed for urban environment where vehicles are equipped with GPS devices and wireless devices for continuous tracking of vehicles. The goal of routing protocols is to select an optimal path with minimum cost. Due to dynamic behavior of VANET network its topology keep on changing. So as to assure the messages are delivered from source to destination time evolving networks should be applied.

Vehicular ad-hoc routing is classified as Unicast, Multicast and Broadcast routing. Unicast routing protocols is useful in vehicle to vehicle communication. Multicast routing is used to disseminate the message to specified area and effective in conditions like traffic jam and accidents. Broadcast routing is

useful to communicate with all the nodes in its transmission range and helpful in conditions like bad road and whether conditions.

In this paper, we focus mainly on unicast routing protocols. The routing protocols for VANET are classified in following categories

3.1 Topology Based Routing

This routing protocol uses topology that exists in the network to forward packets. They are Proactive and Reactive routing protocols.

3.1.1 Proactive Routing Protocol

These kind of routing protocol stores routing information like next forwarding hop. Periodic exchange of topology of network is exchanged among nodes to maintain the paths between any pair of nodes even though some of paths are never used.. A table is then constructed within a node such that each entry in the table indicates the next hop node toward a certain destination. They do not have initial route discovery delay but consumes lot of bandwidth for periodic updates of topology. Various types of Proactive Routing protocols are DSDV, OLSR, FSR, CGSR, WRP.

Advantage
- In these types of protocols, there is no route discovery since the destination route is stored in background.
- They have the best end to-end delay and packets delivery fraction but at the cost of higher routing load.[Comparative Study of Reactive and Proactive Routing]

Disadvantage
- It provides low latency for real time applications.

3.2.2 Reactive Routing Protocol

It is an on demand routing protocol which opens a route only when a node wants to communicate with another node. It maintains the information of only those nodes that are currently being used by network. Thus, reducing burden on network. These types of protocols have a route discovery phase where query packets are flooded into the network in search of a path. The phase completes when a route is found. Reactive routing protocols are applicable to the large size of the mobile ad-hoc networks which are highly mobility and frequent topology changes. AODV, TORA, PGB, DSR are some of reactive protocols.

Advantage
- It minimizes the number of broadcasts since it broadcasts only when route is needed.

Disadvantage
- Excessive flooding can disrupt the network.
- Route finding latency is high.

3.2 Position Based Routing/ Geographic Routing

It is a routing technique in which a node makes its routing decision with the help of information received from GPS device. It sends packet from source to destination based on the geographic position of vehicle instead of using network address. It does not maintain any routing table or exchange link state information with another node in network. In this each node knows its position and its neighbor's position and uses that information to make routing decision. When the source need to send a packet, it stores the destination address in the packet header which will help in forwarding the packet to the destination without needs to route discovery, route maintenance, or even awareness of the network topology. It can be classified as Position based greedy V2V protocols, Delay Tolerant Protocols.

Advantage
- It maintains no information about nodes so it reduces the overhead on network.
- High scalability.

Disadvantage
- Dependent on GPS service.

3.2.1 Position Based Greedy V2V Protocols

GPSR makes the greedy forwarding decision based only on the information about immediate neighbors of a node .when packet reaches a region where greedy forwarding is impossible , the algorithm recovers by routing around the perimeter of the region. There some problem occur with protocol. To solve this problem in this paper we used GSPR-MV protocol.

GSPR-MV
This is called greedy simplified perimeter routing with moving vector. It includes:
- Improved GPSR-MV
- Simplified of perimeter forwarding

In improved GPSR-MV the mobile nodes makes packet forwarding decision directly according to itself, the neighboring nodes and location information of the destination node. There are two kinds of: Greedy and perimeter mode.

In Simplified perimeter forwarding, in the GPSR protocol the nodes listen to transmitted data packets and the no. of experienced hops from the source node to the current node after that the data packet transmission the redundant paths can be avoided and this makes the routes more concise and robust.

3.2.2 Delay Tolerant Protocols

Due to high dynamic topology of VANET frequent disconnection among vehicles is common. It uses carry and forward strategy. It allows nodes to store information when there is no contact between other nodes and it can carry on with that information until it meets with other nodes.

3.3 Cluster based Routing Protocol

This type of routing protocol divides the network into cluster with similar characteristics like topology, speed, velocity etc. Cluster of vehicles is formed and each cluster has a cluster head which is responsible for inter and intra cluster communications. For intra cluster communication, a virtual link is created between nodes for communication and for inter cluster communication cluster communicates through cluster head. Various Clusters based routing protocols are HCB, CBR, COIN, TIBCRPH, CBLR, CBDRP etc.

Advantage
- It provides good scalability for large networks.

Disadvantages
- Due to dynamic topology of VANET the delay and overhead of maintaining these clusters is large.

3.4 Broadcast Based Routing

The simplest way to implement broadcasting is flooding. It is a flooding based routing protocol in which the message is flooded to all nodes in network except form where the message came. It is used by VANET for sharing information like its position, traffic, weather and emergency, road conditions among vehicles to maximum nodes possible. Flooding guarantees that the message will eventually reach all the nodes i.e. vehicles in the network. But in a large network, it causes exponential increase in message transmission resulting in collision so it increases the overhead and decreases the performance of network. Broadcast routing protocols are BROADCOMM, UMB, V-TRADE, DV-CAST, EAEP, SRB, PBSM, PGB, DECA, POCA etc.

Advantage
- Information reaches every node in much less time.

Disadvantage
- Many nodes receive duplicate packets.
- Reduces the network bandwidth.

3.5 Geocast Based Routing Protocol

Geocast routing is a location-based multicast routing. It works by delivering the packet from source to destination within a specified geographical area. The selected area for transmission is called Zone of Relevance or ZOR. The main idea is that sender node need not to packet to nodes beyond the ZOR. Direct flooding strategy is used to reduce the amount of overhead and network congestion when packets are flooded. The various Geocast routing protocols are IVG, DG-CASTOR, DRG.

Advantage
- Reduce network congestion.
- Ensure packet delivery.

Disadvantage
- It requires more latency and control overhead to create routes

Conclusion

In this paper, architecture of VANET and advantages and disadvantages of VANET Routing protocols are discussed. There are number of schemes for handling routing and information dissemination but there are few which deal with safety requirements due to overhead in discovering and maintaining routes. To evaluate the performance of various protocols in VANET can be evaluated based on various performance parameters. Routing vehicle safety communications remain a challenging task.

4. REFERENCES

[1] "Survey of Routing Protocols in Vehicular Ad Hoc Networks," Kevin C. Lee, Uichin Lee, Mario Gerla, Advances in Vehicular Ad-Hoc Networks: Developments and Challenges, IGI Global, Oct, 2009.

[2] "Evaluation of Proactive, Reactive and Hybrid Ad hoc Routing Protocol for various Battery models in VANET using Qualnet" Manish Sharma, Gurpadam Singh, International Journal of Smart Sensors and Ad Hoc Networks (IJSSAN) ISSN No. 2248-9738 (Print) Volume-1, Issue-2, 2011.

[3] "A Comparative Study of Routing Protocol in Vehicular Ad Hoc Network" Pooja Duddalwar, Atul Deshmukh, S. S. Dorle, International Journal of Emerging Technology and Advanced Engineering ISSN 2250-2459, Volume 2, Issue 3, March 2012.

[4] "A Comparative Study of Various Routing Protocols in VANET" Rakesh Kumar, Mayank Dave International Journal of Computer Science Issues, Vol. 8, Issue 4, No 1, July 2011

[5] "A Survey on Routing Mechanism and Techniques in Vehicle to Vehicle Communication (VANET)" Mr. Yugal Kumar, Mr. Pradeep Kumar and Mr. Akash Kadian, International Journal of

Computer Science & Engineering Survey (IJCSES) Vol.2, No.1, Feb 2011.

[6] "Comparative and Behavioral Study of Various Routing Protocols in VANET" Udit Agarwal, Monika Saxena, International Journal of Advanced Research in Computer Science and Software Engineering Volume 3, Issue 10, October 2013.

[7] ''A study on the feasibility of mobile gateways for vehicular ad-hoc networks,'' V. Namboodiri, M. Agarwal, and L. Gao, in proceedings of the First International Workshop on Vehicular Ad Hoc Networks, pp. 66–75, 2004.

[8] "Performance analysis of AODV &GPSR routing protocol in VANET" Neeraj Sharma, Jawahar Thakur International Journal of Computer Science & Engineering Technology (IJCSET) Vol. 4 No. 02 Feb 2013.

[9] "PERFORMANCE EVALUATION OF REALISTIC VANET USING TRAFFIC LIGHT SCENARIO" Nidhi ,D.K Lobiyal, International Journal of Wireless & Mobile Networks (IJWMN) Vol. 4, No. 1, February 2012.

[10] "PERFORMANCE COMPARISON OF POSITION-BASED ROUTING PROTOCOLS IN VEHICLE-TOVEHICLE (V2V) COMMUNICATION" RAM SHRINGAR RAW, SANJOY DAS, International Journal of Engineering Science and Technology (IJEST) Vol. 3 No. 1 Jan 2011.

.

[11] "A Survey on Routing Protocols and its Issues in VANET" Jagadeesh Kakarla, S Siva Sathya, B Govinda Laxmi, Ramesh Babu B International Journal of Computer Applications (0975 – 8887)
Volume 28– No.4, August 2011.

[12] "Survey on security challenges in VANET" Anup Dhamgaye, Nekita Chavhan, International Journal of Computer Science and Network, Vol 2, Issue 1, 2013.

[13] "Survey over VANET Routing Protocols for Vehicle to Vehicle Communication" Bijan Paul, Mohammed J. Islam IOSR Journal of Computer Engineering (IOSRJCE) Volume 7, Issue 5 (Nov-Dec. 2012), PP 01-09.

[14] "VANET Routing Protocols: Pros and Cons" Bijan Paul, Md. Ibrahim, Md. Abu Naser Bikas, International Journal of Computer Applications (0975 – 8887) Volume 20– No.3, April 2011.

[15] Ericson, "Communication and Mobility by Cellular Advanced Radio", ComCar project, www.comcar.de, 2002.

[16] Reichardt D., Miglietta M., Moretti L., Morsink P., and Schulz W.,"CarTALK 2000 – safe and comfortable driving based upon inter-vehicle-communication," in Proc. IEEE IV'02.

[17] "GSPR-MV: a routing protocol based on motion vector for Vanet"Hongyu Tu,Lei Peng, Huiyun li,Falin liu,in IEEE 2014.

Big Data Analytics for the Optimal Delivery of Primary Health Care in China

Muni Kumar N
Department of CSE
SKIT, Srikalahasti
A.P, India

Zongyu Li
Dalian, Liaoning
Province,
China

Jingwen Du
Dalian, Liaoning
Province,
China

Yinghui Zhang
Dalian, Liaoning
Province,
China

Abstract: Heath care is one of the greatest concerns for any country with no exception to China. The country has a large healthcare demand gap due to ageing population, growing urbanization, proliferating lifestyle diseases, rapidly increasing consumer wealth, and the advancement of universal health care insurance. The inefficiencies and inequalities in the public health care access have pushed forward the need for creative thinking and innovative solutions to strengthen the same. Implementation of Electronic Medical Record (EMR), which is the digitized version of medical record which contains notes and documents about patient's identity, examination, therapy, treatments and other services that have been given to the patient. The exponential rise in data in health care has brought lot of challenges in terms of data transfer, storage, computation and analysis. The increase in the number and size of medical records provides new opportunities for identifying patterns and behaviors of the patients and helps the practitioners to make better use of the existing data. This paper discusses the use of Big Data Analytics and Hadoop to reveal the impact of the analytics to render better healthcare services to every citizen of china with optimal cost.

Keywords: Health care in china, Primary Health Care, Data Analytics, Big Data, Analytics in Health Care, Hadoop.

1. INTRODUCTION

China has an exponential increase in the Health care spending. Healthcare Spending are primarily driven by consumer's rapidly increasing incomes and the government's public health care reforms. It is estimated that, the China's health care spending are 5.4 percent of gross domestic product (GDP) in 2013, which is still much lower as compared to Organization for Economic Co-operation and Development (OECD) countries. The country also has a large health care demand gap due to an ageing population, growing urbanization, proliferating lifestyle diseases, rapidly increasing consumer wealth, and the advancement of universal health care insurance. Though, all these elements are driving overall health care market expansion, continued development in health care and better delivery cannot proceed without heavy investment by and strong support from the Chinese government. Public hospitals are the most common and dominant care providers in China. Currently, they rely heavily on medicine charges for funding. In long run, This automatically weakens the public social security system which leads to medical resources allocation inefficiency and prompts patients to switch to private hospitals.

Health strategies costs[1] in China will almost triple to more than $1 trillion annually by 2020, driven by an aging population and government efforts to broaden insurance coverage making it the second-largest healthcare economy after the United States. More than 95% of China's 1.4 billion citizens have access to health insurance. However, benefits are rather limited and there are wide gaps in access to

healthcare between urban centers, with its far-superior hospitals, and rural centers, where living standards are considerably lower. To address these disparities, the Chinese government has been embarking on a variety of health reform initiatives like Regional medical and telemedicine services. Community medical and health service systems will still remain a priority investment area. In addition, closer connections will be established between medical services and medical insurance, management of population, and pension. While health reform is having a positive impact on China's citizens, these initiatives are creating a more difficult environment for multinational companies that want to expand into the China market. More now than ever, IT suppliers will need local partners to compete effectively. With the implementation of Telemedicine services in china, the huge gap in health care provision and quality can be minimized in the rural parts of the country and better health care delivery can be guaranteed with the combination of telemedicine and big data analytics.

To address China's huge irregularities in health care provision and quality, the government is anticipated to increase spending in rural areas for the establishment of clinic with sophisticated equipments, supply of sufficient drugs and for the patient insurance. Research studies reveal that, Ninety-six (96%) percent of China's population is covered by some form of medical insurance either through national health insurance or a rural co-operative medical insurance system. However, the coverage is very less with low premium contributions and patients continuing to pay large outpatient fees.

Private and foreign hospitals in china can bring leading medical technologies and advanced management, clinical practices, and service models to China. Also, Chinese officials have encouraged the development of private health insurance as a supplement to the public scheme. In fact, some local governments have formed partnerships with private companies to manage their public insurance plans. The expansion of private hospitals and private insurance should generate growth opportunities to related industries along the health care value chain, such as senior care, medical tourism, health management, medical devices, mobile health (mHealth) and other health information technology (HIT). The downside is that most of these new health services are not currently within the scope of social insurance coverage; the advancement of commercial insurance in China could help to mitigate these difficulties.

This paper is organized as follows: section II discusses about the Literature Review with focus on the Health care delivery in china, section III discusses about big data and its characteristics, section IV focuses on various big data technologies that can be applied to health care, section V deals with the application of big data analytics to health care and section VI concludes the work.

2. LITERATURE REVIEW

This literature aims to highlight the range of medical facilities provided by the health care in China. These literatures have emphasized upon the fact that Chinese hospitals and medical centers are categorized into three major types such as primary, Secondary or Tertiary[2]. According to Hou, these facilities are involved in the provision of preventive care, nominal health care and rehabilitation examination of different types of patients [3]. In the process of discussing the primary care quality among health care structures in Tibet, China, Wenhua Wang, Shi and Yin have highlighted that still differences in township and country hospitals is visible for better primary care[4]. The author reported that primary care in China was aimed to provide first contact and inclusive health services by lowering cost to different geographic areas but currently due to structural problems, the primary care system is facing with imbalanced situations. In contrary, Rohde, Cousens and Chopra have identified that current compressive primary care system in China based on dependable referral system is better than the countries with selective primary care [5].

On the other hand, "*China allocates 5.6% to its total budget to health care sector. Further, China spends 35% of its allotted budget on the primary health and 65% of its budget on the secondary and tertiary health care*". The study gave credit to the new china health reforms in bringing a remarkable transformation in the primary health and balanced approach. Likewise, researches have further pointed out that secondary and tertiary cares are provided in public hospitals, institutions, and specialist outpatient clinics (SOPC). These care facilities are involved in surgery and emergency services.

In order to investigate further that how health care is currently being provided to people in China, past authors have also covered information about the types of insurance available for patients in China. According to The State Council Information Office of the People's Republic of China, Chinese health care system is based on multi-layered insurance system. The system offers range of supplementary medical insurance and commercial health insurance to the urban and rural populations for providing security and opportunities to avail different types of cares. In a research conducted by Hu and Ljungwall [6], authors have mentioned different names of the medical insurance systems introduced in China for helping their population. These include urban employee basic medical insurance (UEBMI), urban resident basic medical insurance (URBMI), rural cooperative medical system (RCMS). The report further highlighted that UEBMI is funded through the mix of the social pooling and individual medical savings while funds for URBMI is gathered from the households and only up to certain extent, government provides subsidies for it. Likewise, RCMS is also the outcome of combined individual payments, financial assistance from communal businesses and state-sponsored subsidies. The point directs that Chinese government does not bear the entire burden of health care services at present.

Additionally, it can be examined from the overview of past literature that at present preventive care is the vital component of Chinese health system but current reforms do not consider it effectively. The government has introduced anti-smoking initiatives for dealing with the diseases caused through drugs intake. Likewise, health-ageing programs, established of community support centre for cancer patients and immunization are also visible in the past studies. Xin has further added in the discussion by stating that reimbursements are also vital source of the preventive care in the China to increase affordability of patients[6]. Similarly, another example of preventive care in China can be examined from the HIV/STD prevention programs initiated by the health care system for increasing communication and awareness about the sex education among the population [7].

Thus, based on the findings gathered from different studies in the literature review, it can be examined that current provisions and health care facilities in China are multi-dimensional, which has taken account of different types of patients (primary, secondary, tertiary, preventive and catastrophic care patients. These services are to reformed also cover both financial and non-financial aspects.

3. BIG DATA AND ITS CHARACTERISTICS

Big data is the amount of data created in the world up to and including 2005 is now created every 2-days. Big data is a platform for importing, storing and analyzing data to uncover information not previously known. This explosion of the data, changing the way people think about everything. From the cutting edge scientific research to the monetization of social

media and exchanging the way people think about healthcare analytics too. However, the health care has not kept pace with Big Data

Historically, data was being generated and accumulated by workers. In other words, employees of companies were entering data into a computer system. But, then the things are evolved with the Internet and now users could generate data on their own. For eg. In the websites like Face book, Google, Amazon etc. all the users are signing up and entering the data themselves, which is scalable and it is much larger than the first in magnitude. Regarding scalability, it is scaled up from just employees entering the data to users entering their own data, so all of a sudden the amount of data being generated and accumulated was very higher than it was historically. Now, the third level is, the machines are accumulating data. The buildings all over the cities are full of monitors that are monitoring humidity and temperature, there are satellites around the earth that are monitoring the earth 24-hours a day taking pictures accumulating data. Once machines started accumulating data, the data has become much voluminous. Back in the good old days, people used to use relational databases to process the data and bring it to the processor that the CPU processes the data. But, now there is so much data that overwhelms the power of CPU and the CPU cant process as the data is very huge. So, now what the people are doing is, they are bringing multiple processors to the data and do parallel processing. So, now the data is being processed in a whole bunch of different places at the same time. Before, the data is being brought to the processor, but now the processor is being brought to the data to processing. In the first case, data is brought to one CPU, But now the we could bring in infinite number of CPU's to infinite number of individual servers and do parallel processing. So, now the data has grown to larger orders of magnitude to higher and now we have a way to process higher magnitude data as well, that the technological shift. Some of the technologies that are allowing this to happen are Hadoop, Map Reduce, NOSQL.

3.1 Definition

Big Data is a collection of large and complex data sets which are difficult to process using common database management tools or traditional data processing applications. According to zdnet.com, "Big data refers to the tools, processes and procedures allowing an organization to create, manipulate and manage very large data sets and storage facilities".

Big data is being generated by everything around us at all times. Every digital process and social media exchange produces it. Systems, sensors and mobile devices transmit it. Big data is arriving from multiple sources at an alarming velocity, volume and variety. To extract meaningful value from big data, you need optimal processing power, analytics capabilities and skills. Big data is changing the way people within organizations work together. It is creating a culture in which business and IT leaders must join forces to realize value from all data. Insights from big data can enable all employees to make better decisions—deepening customer engagement, optimizing operations, preventing threats and

fraud, and capitalizing on new sources of revenue. But escalating demand for insights requires a fundamentally new approach to architecture, tools and practices.

The big data are generated from online transactions, emails, videos, audios, images, click streams, logs, posts, search queries, health records, social networking interactions, science data, sensors, mobile phones and their applications[8].
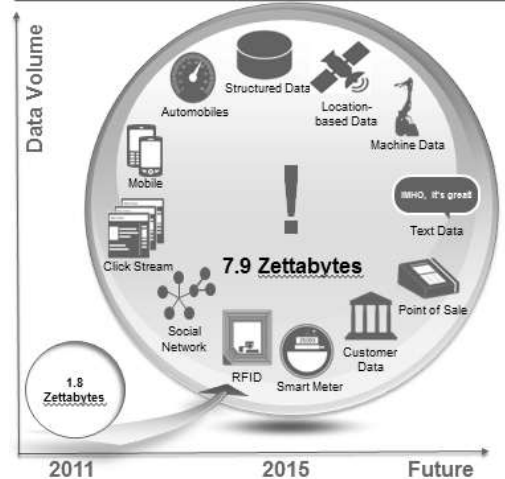

Fig.1 Voluminous Data generated through different sources

Big data is a platform for importing, storing and analyzing data to uncover information not previously known. This explosion of the data, changing the way people think about everything. From the cutting edge scientific research to the monetization of social media and exchanging the way people think about healthcare analytics too. However, the health care has not kept pace with big data.

The large Chinese health care system needs to harness healthcare's "big data" and analyze a complex set of data, including electronic medical records and sensor data. This enables clinicians to access and analyze healthcare big data to ascertain quality, determine best practice, assess treatment strategies and identify patients at risk. The promises and potential of big data in transforming digital government services, governments, and the interaction between governments, citizens, and the business sector, are substantial. From "smart" government to transformational government, Big Data can foster collaboration; create real-time solutions to challenges in agriculture, health, transportation, and more; and usher in a new era of policy- and decision-making. Big Data raise a large number of information management issues, primarily in the areas of privacy, security, accuracy, and archiving, spanning major issues such as personally identifiable information, security of government data and information, and the accuracy of publicly available data. By fostering collaborations and economic development through private-public partnerships, government agencies appear to be tacitly endorsing the privacy, security, and other policies employed by those private sector entities.

3.2 Characteristics of Big Data

Big data generally refers to the social network data from the micro-blogging sites like Twitter, LinkedIn and social media platforms like Face book, Traditional enterprise including transactional data, web store transactions etc. and machine generated / sensor data like call data records, smart meters, manufacturing sensors, trading systems, traffic data, air data etc. which keeps on increasing without the human intervention. Big data is not only driven by the exponential growth of data but also by changing user behavior and globalization. Globalization provides competition among the participants in the market. As a result, organizations are constantly looking for opportunities to increase their competitive advantage by using better analytical models.

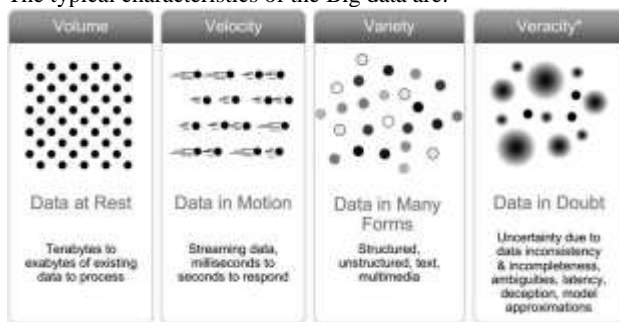The typical characteristics of the Big data are:

Fig 2. Characteristics of Big data

3.2.1 Volume: The amount of data being generated by various health care monitoring systems connected to patient over a period of time will be very huge in size. This voluminous data will become difficult for the conventional systems to handle and manage.

3.2.2 Velocity: Velocity is the speed at which the data is being generated like streamed data from various smart devices into social media and also camera streamed data which stores the data in motion from huge number of closed circuit cameras. In health care systems, the data will be generated by the sensors connected to the patients and will be generated with a high velocity.

3.2.3 Variety: The Variety refers to the various formats of data like structured, semi-structured or un-structured. The data that will be generated at the health care organization will be in different forms like text, documents, emails, audio, video, text messages, graphs, diagnostic reports etc.

3.2.4 Veracity: Veracity refers to the biases, noise and abnormality in data. Veracity in data analysis is the biggest challenge when compared to other characteristics like volume and velocity.

3.2.5 Value: Value refers to what type of value does the big data generates. This characteristic deals with the various ways of presenting results values like statistical, correlation, hypothetical etc.
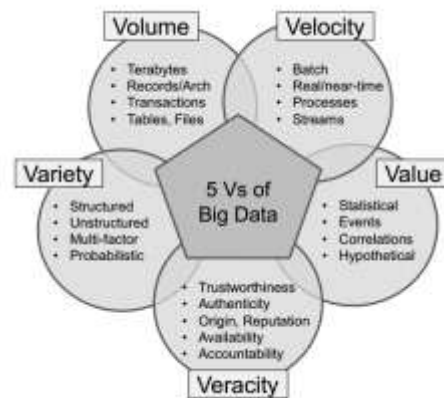

Fig.3: Characteristic of Big Data. (5 Vs of Big Data)

4. BIG DATA TECHNOLOGIES

4.1 Hadoop: Apache Hadoop [9] is an open source software project that enables the distributed processing of large data sets across clusters of commodity servers. It is designed to scale up from a single server to thousands of machines, with a very high degree of fault tolerance. Rather than relying on high-end hardware, the resiliency of these clusters comes from the software's ability to detect and handle failures at the application layer.

Hadoop is an open source platform that is developed under General Public License (GPL) and it is developed by the developers across the world and it is free to use. It organizes the parallel processing and this is the software that allows parallel processing to happen and the second is called MapReduce. The Hadoop and Map Reduce are working together for the processing and analytics of Big Data. The companies like Google, Amazon and many others are using these processes together to discover unbelievable insights that is hidden inside all of that Big data.

Components of HADOOP

- YARN - Yet Another Resource Negotiator (YARN) assigns CPU, memory, and storage to applications running on a Hadoop cluster.The first generation of Hadoop could only run MapReduce applications. YARN enables other application frameworks (like Spark) to run on Hadoop as well, which opens up a wealth of possibilities.

- HDFS - Hadoop Distributed File System (HDFS) is a file system that spans all the nodes in a Hadoop cluster for data storage. It links together the file systems on many local nodes to make them into one big file system.

Features of Hadoop

- Scalable– New nodes can be added as needed, and added without needing to change data formats, how data is loaded, how jobs are written, or the applications on top.

- Cost effective–Hadoop brings massively parallel computing to commodity servers. The result is a

sizeable decrease in the cost per terabyte of storage, which in turn makes it affordable to model all your data.

- Flexible–Hadoop is schema-less, and can absorb any type of data, structured or not, from any number of sources. Data from multiple sources can be joined and aggregated in arbitrary ways enabling deeper analyses than any one system can provide.
- Fault tolerant– When you lose a node, the system redirects work to another location of the data and continues processing without missing a fright beat.

4.2 Map Reduce

Map Reduce [10] is a software framework for easily writing applications which process vast amounts of data (multi-terabyte data-sets) in-parallel on large clusters (thousands of nodes) of commodity hardware in a reliable, fault-tolerant manner. A Map Reduce job usually splits the input data-set into independent chunks which are processed by the map tasks in a completely parallel manner. The framework sorts the outputs of the maps, which are then input to the reduce tasks. Typically both the input and the output of the job are stored in a file-system. The framework takes care of scheduling tasks, monitoring them and re-executes the failed tasks.

Map Reduce is a programming model and an associated implementation for processing and generating large data sets with a parallel, distributed algorithm on a cluster. Typically the compute nodes and the storage nodes are the same, that is, the Map Reduce framework and the Hadoop Distributed File System are running on the same set of nodes. This configuration allows the framework to effectively schedule tasks on the nodes where data is already present, resulting in very high aggregate bandwidth across the cluster.

The Map Reduce framework consists of a single master Job Tracker and one slave Task Tracker per cluster-node. The master is responsible for scheduling the jobs' component tasks on the slaves, monitoring them and re-executing the failed tasks. The slaves execute the tasks as directed by the master.

4.3 Hadoop Distributed File System (HDFS)

The Hadoop Distributed File System (HDFS) is a distributed file system designed to run on commodity hardware. It has many similarities with existing distributed file systems. However, the differences from other distributed file systems are significant. HDFS is highly fault-tolerant and is designed to be deployed on low-cost hardware. HDFS provides high throughput access to application data and is suitable for applications that have large data sets.

Name nodes and Data nodes: HDFS has a master/slave architecture. An HDFS cluster consists of a single Name node, a master server that manages the file system namespace and regulates access to files by clients. In addition, there are a number of Data nodes, usually one per node in the cluster, which manage storage attached to the nodes that they run on.

HDFS exposes a file system namespace and allows user data to be stored in files. Internally, a file is split into one or more blocks and these blocks are stored in a set of Data nodes. The Name node executes file system namespace operations like opening, closing, and renaming files and directories. It also determines the mapping of blocks to Data nodes. The Data nodes are responsible for serving read and write requests from the file system's clients. The Data nodes also perform block creation, deletion, and replication upon instruction from the Name node.
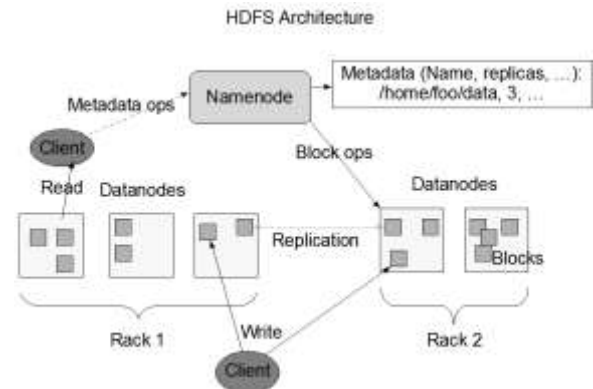

Fig 4: HDFS Architecture

The NameNode and DataNode are the pieces of software designed to run on commodity machines. These machines typically run a GNU/Linux operating system (OS). HDFS is built using the Java language; any machine that supports Java can run the Name node or the Data node software. Usage of the highly portable Java language means that HDFS can be deployed on a wide range of machines. A typical deployment has a dedicated machine that runs only the Name node software. Each of the other machines in the cluster runs one instance of the Data node software. The existence of a single Name node in a cluster greatly simplifies the architecture of the system. The Name node is the arbitrator and repository for all HDFS metadata. The system is designed in such a way that user data never flows through the Name node.

5. CRITICAL ISSUES IN HEALTH CARE

The following are the critical issues to be considered to deliver optimal health care in rural parts of the china and bridge the gap between quality and affordability in government hospitals. These ideas will enable us to access the services on par with the private super specialty hospitals. Further, the implementation of these issues will provide cheaper, better and easier health care facilities to the citizens of India.

1. **e-Health File:** The creation of a e-Health care file for each patient, where all health care providers and patients themselves were able to submit information (with the consent of the patient). Both subjective data, symptom diaries, lab data, image diagnostics, pathology reports etc., could be filed. To overcome the information overload from the massive amounts of data, Big Data Analytics could be employed for the processing of the

data and obtain the desired results with great accuracy in reasonable time.

2. **Creating awareness with chronic diseases:** The system must identify and create awareness among the people with the common chronic diseases at particular areas, through which we can prevent diseases. These chronic diseases are responsible for the 75% of health care spending due to lack of awareness and prior care.

3. **e-Prescribe:** Paper based prescriptions are archaic and lead to several miseries each year due to errors in prescription. But if every doctor is provided with an electronic prescription system, it would improve safety by making prescriptions easier to read and providing instant checks on drug interactions, dosages, and a patient's medication history.

4. **Electronic Medical Records:** Medical Experts agree that electronic medical records (EMRs) are a must for the better health care in India. But, at present only few hospital are maintaining EMR's, mainly because of cost, privacy issues, and the lack of one compatible, easy-to-use infrastructure.

5. **Stop Unnecessary Treatments:** Doctors should avoid trial and error type of medication. The problem must be examined thoroughly by performing the required diagnostic tests during the preliminary days of disease. The right treatment should be suggested at the first visit only which avoids the disease to become more critical. Most of the issues are arising with the misdiagnosis and wrong treatment during the early stages.

6. **Reduce Infant Mortality:** The Infant Mortality rate in our country is substantially large compared to other countries. Though, the government has several schemes for the pregnant women, those were not yielding better results due to the lack of proper medical care. If the proper care is taken towards the pregnant women, definitely the new-born baby will be healthy in all aspects to make India healthy.

7. **Tele Medicine:** Doctors can often diagnose or prescribe without seeing the patient. The patient has to physically appear before the nearby health center, where the nurses or health workers will diagnose at first level, note the symptoms and informs the high level specialist doctors about the case. After examining the reports, the specialist doctor suggests the treatment through health worker which reduces costs and creates satisfaction by virtual communication of patients and doctors to discuss medication changes and test results through an online system.

5.1 Tele-Medicine & EMR

Telemedicine can be defined as the use of electronic communication technology to exchange patient information and provision of health care services at remote locations. Global Telemedicine has gone far beyond providing health care services alone. It is now being extensively used also for education, research and management of data. It is, however, paradoxical that despite India's strength in information technology, the use of telemedicine is still at a fairly nascent stage especially in the public health sector.

Use of Information and Communication Technology (ICT) in health can be broadly in four areas viz. Education, Research, Referral, and Management of Data.

- **Health and Education** : When ICT is applied to medical education, it is possible to make high quality education available pan India seamlessly.

- **Hospital Management System :** It is common experience that work places in the health systems are not adequately governed especially in remote areas. The management of medical colleges and district hospitals can be strengthened significantly with the application of Hospital Management System.

- **Health Research :** ICT can potentially transform the medical scene in India, by bringing about a sea-change in medical research. From traditional clinical research to the modern synthetic biology-based research, the opportunity is immense. Work on problems such as Cancer prevention, screening, diagnosis, and therapy can benefit from inter disciplinary cooperation. Medical fraternity has availed such benefits when MRI and Nuclear Imaging was integrated into medicine a few decades ago.

- **Managing Health Data :** Electronic Medical Records (EMR), is a fundamental pre-requisite in using ICT seamlessly in healthcare. While EMR is available in several forms, size, shape, and format, the medical community with a specific mandate should standardize EMR, create and establish ICT platforms for using EMR based systems, for universal benefits.

6. ROLE OF BIG DATA IN HEALTH CARE

We live in the age of big data. The amount of data created in the world up to and including 2005 is now created every two days. Big data is a platform for importing, storing and analyzing data to uncover information not previously known. This explosion of the data changing the way people think about everything. From the cutting edge scientific research to the monetization of social media and exchanging the way people think about healthcare analytics too. However, the health care has not kept pace with big data. Big Data Healthcare is the drive to capitalize on growing patient and health system data availability to generate healthcare innovation. By making smart use of the ever-increasing amount of data available, we find new insights by re-examining the data or combining it with other information. In healthcare this means not just mining patient records, medical images, diagnostic reports etc., for insights, diagnoses and decision support device, but also continuous analysis of the data streams produced for and by every patient in a hospital, at home and even while on the move via mobile devices [11].

Even today the majority of health care analytics is performed by doing monthly data refreshes in relational databases that produce pre-processed reports. A fair gap is often missing lab

test is often 45 days old, as the data flow move from batched data fields to real time fields from transactional systems and streaming data from analytical modeling devices. This old model of analytics will fail. Analysis will need to be done on that spot moment not in the pre-processed form. Data refreshes need to be done in real-time not once in a month. The data analysis tools of today are likely yellow pages phone book in the era of Internet Search Engine. They are becoming more obsolete with each passing day. The traditional health care analytic tools are built on tools developed by IBM in 1970, more than 40 years ago.

If all the three parties (payer, provider, pharmaceutical company) [12] work collaboratively and share data/insight, disease management programs will become cost-effective and deliver improved patient outcomes at a scale that will further optimize overall health-care cost structure.

The term "e-health" [13] defined by WHO: "a new term used to describe the combined use of electronic communication and information technology in the health sector". e-health is the main driver for three significant changes within the health care environment[13]:

1. Patients to become better informed
2. Patients to become more active and empowered in their health care
3. Healthcare to become more efficient.

It is referred in the Cognizant 20-20 insights by Cognizant, [14] Big data solutions attempt to cost-effectively solve the challenges of large and fast-growing data volumes realize its potential analytical value. For instance, trend analytics allow you to figure out what happened, while root cause and predictive analytics enable understanding of why it happened and what it is likely to happen in future. All healthcare constituents – patients, payers, providers, groups, researchers, governments etc. – will be impacted by big data, which can predict how these players are likely to behave, encourage desirable behavior. These applications of big data can be tested, refined and optimized quickly and inexpensively and will radically change healthcare delivery and research. The healthcare domain has been an easy target for people who seek easy money by using fraud methods [15]. Healthcare fraud is expected to continue to rise as people live longer. The white paper by trend analytics [15] reveals that healthcare fraud prevention has resulted savings of nearly $4.1 billion in 2011. A big data platform has ability to sift through a huge amount of historical data in relatively shorter amount of time, so that the business transactions can use fraud detection on real time. Though, the big data analytics in healthcare plays a crucial role to provide better health care services, provide analysis on the historical data to uncover hidden information, the big data analytics has the challenges like Heterogeneity and Incompleteness of data, scale, timeliness, privacy and Human Collaboration. The future research is all about to overcome the challenges and use big data analytics in healthcare to uncover the knowledge from the raw unstructured data.

Opportunities for big data in health care

Big Data has a large potential to contribute in many areas of the health care Industry. At the moment, there are some good initiatives, but this is not enough to keep up with the demand of health care service and the rising costs.
The opportunities for Big Data in health care are:

1. Electronic Health records (EMR/EHR) which serve the customer/patient
2. Structuring data and information for service optimization
3. Accurate information about patients can reduce mistakes
4. Cost optimization through efficiency of new e-health services
5. Increased customer satisfaction (better e-health services)
6. Analysis of big datasets for R&D purposes

7. CONCLUSION

Big Data Analytics in Hadoop's implementation provides systematic way for achieving better outcomes like availability and affordability in healthcare service to all population. In this paper we discussed about the implementation of Electronic Medical Records (EMRs) in China. Further, this paper discusses about various ways of transforming health records (EHRs) to useful and analyzed result, this analysis will make the patient to understand the complications to occur and thereby the patient will be cautious towards his health status or health behavior. The goal of this research deals with the study and application of big data analytics to health care industry.

8. FUTURE WORK

In the future, the present research work will be extended with the implementation of Hadoop, HDFS and MapReduce to handle huge amount of semi-structured/ un-structured health care data and also provide the real analytics through which the patients and doctors will enjoy the real benefit of Hadoop and Big Data Analytics.

9. ACKNOWLEDGMENTS

10. REFERENCES
[1] Shi, Leiyu and Douglas A. Singh. *The Nation's Health.* London: Jones & Bartlett Learning, 2011.

[2] Hou, Thomas. The Chinese Primary Care System: Its Evolution, Challenges and Legal Aspects of Reform. Research Paper. Pennsylvania: University of Pennsylvania, 2009.

[3] Wenhua Wang, et al. "Primary Care Quality among Different Health Care Structures in Tibet, China." *BioMed Research International* (2015): 1-8.

[4] Rohde, J, et al. "30 years after Alma-Ata: has primary health care worked in countries?" *The Lancet, 372(9642)* (2008): 950-961.

[5] Hu, Shanlian and Christer Ljungwall. *China's Healthcare System - – Overview and Quality Improvements.* Ostersund: Swedish Agency for Growth Policy Analysis, 2013.

[6] Xin, Li. "The Effect of China's New Cooperative Medical Scheme on Rural Utilization of Preventive Medical Care and Rural Households' Health Status." *Master of Pacific International Affairs* (2011): 1-20.

[7] Lieber, Eli, et al. Sociocultural Contexts and Communication about Sex in China: Informing HIV/STD Prevention Programs. Los Angeles: UCLA/NPI , 2014.

[8] Sagiroglu, S.; Sinanc, D., "Big data: A review," Collaboration Technologies and Systems (CTS), 2013 International Conference on , vol., no., pp.42,47, 20-24 May 2013

[9] http://www-01.ibm.com/software/data/infosphere/hadoop/accessed on 30-08-15

[10] http://hadoop.apache.org/docs/r1.2.1/mapred_tutorial.html accessed on 30-08-15

[11] Edwin Morley-Fletcher, "Big Data Healthcare, An overview of the challenges in data intensive health care", Discussion Paper.

[12] SoumendraMohanty, MadhuJagadeesh, and HarshaSrivatsa,"Big Data Imperatives", Apress publications. 2013

[13] Manuel Zwicker, Juergen Seitz and NilminiWickramasinghe, "Identifying Critical Issues for Developing Successful E-Health Solutions", Proceedings of the Pacific Asia Conference on Information Systems (PACIS 2012). Paper 33. http://aisel.aisnet.org/pacis2012/33

[14] Bill Hamilton , "Big Data is the Future of Healthcare", Cognizant 20-20 Insights, September 2012.

[15] VenkatareddyKonasani, MukulBiswas and Praveen Krishnan Koleth, " Healthcare Fraud Management using Big Data Analytics", A Whitepaper by Trendwise Analytics.

[16] Fraud Detection Using Data Analytics in the Healthcare Industry - Discussion white paper

Duplicate Code Detection using Control Statements

Sudhamani M
Department of Computer Science
University of Mysore
Mysore 570006, India

Lalitha Rangarajan
Department of Computer Science
University of Mysore
Mysore 570006, India

Abstract: Code clone detection is an important area of research as reusability is a key factor in software evolution. Duplicate code degrades the design and structure of software and software qualities like readability, changeability, maintainability. Code clone increases the maintenance cost as incorrect changes in copied code may lead to more errors. In this paper we address structural code similarity detection and propose new methods to detect structural clones using structure of control statements. By structure we mean order of control statements used in the source code. We have considered two orders of control structures: (i) Sequence of control statements as it appears (ii) Execution flow of control statements.

Keywords: Control statements; Control structure; Execution flow; Similarity value; Structural similarity.

1. INTRODUCTION

Duplicate codes are identical or similar code fragments present in software program. Two code fragments are similar if these code segments are similar in their structure of control statements and similar control flow between control lines [1, 15].

Different types of code clones are [15]

Type 1: Exact similar code fragments except white space and comments as shown in below example.

Ex 1:

Segment 1:
```
if(n>0)
{
 n=n*1;  //multiply by plus 1
}
else
n=n*-1; // multiply by minus 1
```

Segment 2:
```
if ( n > 0 )
{
 n = n * 1;  //multiply by +1
}
else
n = n * -1;  // multiply by -1
```

Type 2: Syntactic similar code fragments except change in variable, literal and function names.

Ex 2:

Segment 1:
```
if (n>0)
{
 n=n*1;  //multiply by plus 1
}
else
n=n*-1; // multiply by minus 1
```

Segment 2:
```
if ( m > 0 )
{
 m = m * 1;  //multiply by +1
}
```
```
else
m = m * -1;  // multiply by -1
```

Type 3: Similar code fragments with slight modifications like reordering/addition/deletion of some statements from already existing or copied code fragments.

Segment 1: if (n > 0)
```
    {
     n=n*1;  //multiply by plus 1
    }
    else
    n=n*-1;  // multiply by minus 1
```

Segment 2: if (n > 0)
```
    {
     n=n*1;  //multiply by plus 1
    }
    else
    n=n*-1;  // multiply by minus 1
    x=5;   //newly added statement
```
In the above example a new statement x=5 is added.

Type 4: Functionally similar code fragments. Below example explains recursive and non recursive way of finding factorial of n. (same program implemented in two ways).

Ex:

Segment 1: int i, j=1, n;
```
for (i=1; i<=n; i++)
j=j*i;
```
segment 2:
```
int fact(int n)
{
if (n == 0) return 1 ;
else return n * fact(n-1) ;
}
```

Output of program depends on the execution flow of effective source lines. Execution flow of source lines depends on the control lines used in the program. Control lines considered here are iterative statements (for, while and do-while), conditional statements (if, if-else and switch-case), and

```
Sorting program 1                    Sorting program 2                 Sorting program 3
main ()                              Sort()                            Sort()
{                                    {                                 {
int a[30],n,  i, j, temp;            for (i=1; i <=n; i++)             for (j=1; j<=n; j++)
printf ("\n Enter the size of an array");  for (j=1; j<=n; j++)        if (a[j] >a[j+1])
scanf ("% d",&n);                    if  (a[j] >a[j+1])                {
printf ("\n Enter the elements to    {                                temp = a[j];
sort");                              temp = a[j];                      a[j] = a[j+1];
For (i=1; i<=n; i++)                 a[j] = a[J+1];                    a[j+1] = temp;
scanf ("% d",&a[i]);                 a[j+1] = temp;                    }
                                     }                                 }
printf ("numbers to sort are");      }
                                     Print()                           Print()
for(i=1; i<=n; i++)                  {                                 {
{                                    for( i=1; i<=n; i++)              for (i=1; i<=n; i++)
printf ("% d ",a[i]);                printf  ("% d ",a[i]);            Printf ("% d ",a[i]);
}                                                                      }
                                     main () {
For (i=1; i <=n; i++)                int a[30],n, i, j, temp;          main ()
{                                    Printf ("\n Enter the size of an array");   {
for (j=1; j<=n; j++)                 scanf ("% d",&n);                 int a[30],n, i, j, temp;
{                                    printf ("\n Enter the elements to Printf ("\n Enter the size of an
if (a[j] > a[j+1])                   sort");                           array");
{                                    i=1;                              Scanf  ("% d",&n);
temp=a[j];                           while (i <= n)                    Printf ("\n Enter the elements to
a[j]= a[J+1];                        {                                 sort");
a[j+1] = temp;                       Scanf ("% d",&a[i]);              for (i=1; i<=n; i++)
}                                    i++;                              Scanf ("% d",&a[i]);
}                                    }                                 printf ("numbers to sort are");
}                                    Printf ("numbers to sort are");   Print();
printf ("Sorted numbers");           Print();                          for (i=1; i <+n; i++)
for(i=1; i<=n; i++)                  Sort();                           Sort();
printf ("% d ",a[i]);                printf ("Sorted numbers");        printf("Sorted numbers");
}                                    Print();                          Print();
                                     }                                 }
```

Fig 1: Different versions of bubble sort program

function call. Here we propose two approaches to find structural similarity. Approach 1 considers order of control statements present in the code segments and approach 2 depends on the execution flow of control lines in the program. Figure 1 shows three different ways of writing bubble sort program. To find similarity of these programs we compute control structure metrics. Rest of the paper is organized as follows. Section 2 covers key literature, section 3 describes proposed methods and results; section 4 concludes the work with suggestions on possible future work.

2. RELATED WORK

Duplicate code detection mainly consists of two phases where first phase is transformation and second phase is comparison. In transformation phase, source code is transformed in to an Internal Code Format (ICF). Depending on the ICF comparison, match detection techniques are classified as follows [15].

i. **String Based:** In these techniques source code is considered as an arrangement of characters/strings/lines and uses string matching techniques to detect duplicate code [2]. Dup tool compares lexemes on behalf of string match and finds partial match [2, 3, 4]. Ducass et al [5] proposed dynamic matching technique to detect code clones. String based techniques are simple, language independent and detect type I clones [13, 14, 15, 16].

ii. **Token Based:** In token based approach source code is transformed into sequence of tokens using lexer/parser. Then these sequences of tokens are compared to find duplicate code. This technique detects both type I and II clones. Kamiya et al's [5] CC Finder regenerate source file into a set of tokens and device single token from these set of tokens and

uses suffix tree substring matching algorithm to detect code clones. CP Miner uses frequent substring matching algorithm to replicate tokenized statement. SIM correlate the chain of tokens using dynamic programming string alignment technique. Winnowing and JPlag are token based plagiarism detection tools [13, 14, 15, 16].

iii. **Tree Based:** Source text is parsed to obtain Abstract Syntax Tree (AST) or parse tree with appropriate parser. Then tree matching techniques are used to find similar sub trees. This approach efficiently detects type I, type II and type III clones [5, 6]. As AST does not address data flow between controls, it fails to detect type IV clones. Baxter et al's CloneDR find resemblance between programs by matching sub trees of corresponding source program [15].

iv. **Graph Based:** Source program is converted into Program Dependency Graph (PDG) where PDG contains the data flow and control flow information of the program [6]. Then isomorphic sub graph detection algorithms are used to find duplicate code. This technique efficiently identifies all types of clones. However generating PDG and finding isomorphic sub graphs is NP hard [8]. Komondoor and Horowitz PDG-DUP uses program slicing to find isomorphic sub graphs, Krinke uses iterative approach to detect highest comparable sub graphs. GPLAG is graph based plagiarism disclosure tool [11, 16].

v. **Metric Based:** In this technique different metrics are computed for code fragments and these metric values are compared to find duplicate code [9, 10, 11, 12]. AST/PDG representation can be used to calculate metrics like number of nodes, number of control edges present in the graph etc. Other common metrics are number of source lines, number of function calls, number of local and global variables and McCabe's cyclomatic complexity etc. eMetric, Covert and

Moss are metric based tools [15, 16]. Kontogiannis et al. [16] build an abstract pattern matching tool to identify probable matches using Markov models to measures similarity between two programs.

3. PROPOSED METHOD

Here we propose two approaches to find duplicate code. The different stages in the proposed method are preprocessing, metric computation, difference matrix computation and similarity value calculation. Architecture of proposed method is shown in figure 2 and each stage is explained subsequently.

Preprocessing and template conversion
In preprocessing stage extra space and comments are removed and input source program is transformed into its standard intermediate template form. Figure 3 shows the template form of versions of sort program in figure 1. This template is used to compute control structure metrics.
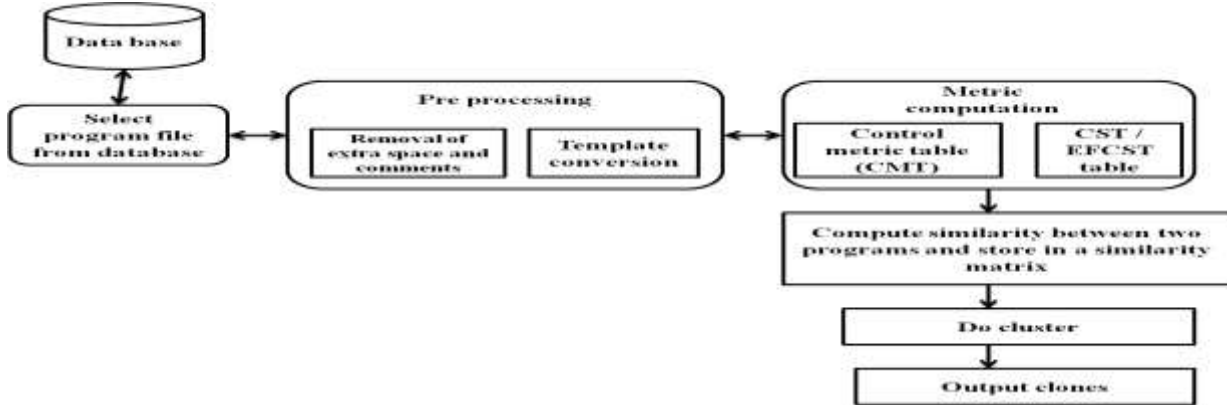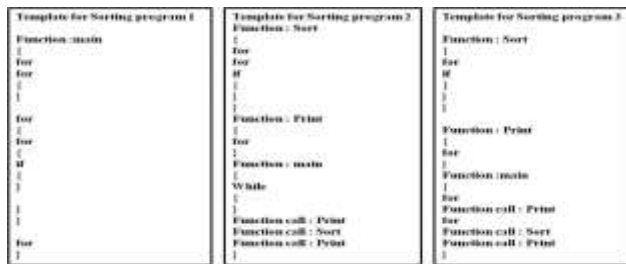

Fig 2: Architecture of proposed method

Fig 3: Templates of sort programs in figure 1

Note that the order / structure of control statements are different across versions. Some versions have function calls and some don't. Yet proposed approaches can detect duplicate to high accuracy.

3.1 Approach 1 – Computation of similarity using Control Structure Tables (CSTs)

Control Structure Table (CST): Control Structure Table contains the information about order of ingrained control lines used in the program [11]. CST of sort program 1 and sort program 2 in figure 1 are shown in table 1 and 2.

Table 1. Control structure table for sort program 1

Sl.No	Type of control statement	Loop	Condition
1	Loop	0	0
2	Loop	0	0
3	Loop	1	1
4	Loop	0	1
5	Condition	0	0
6	Loop	0	0

Table 2. Control structure table for sort program 2

Sl.No	Type of control statement	Loop	Condition
1	Loop	1	1
2	Loop	0	1
3	Condition	0	0
4	Loop	0	0
5	Loop	0	0

Difference Matrix (D) computation: Difference matrix is calculated using two CSTs. Difference matrix calculated from table 1 and 2 are shown in table 3. Difference matrix shows different between all pairs of control statement. Difference matrix (D) is computed from the respective control structure tables. A row of program 1 (corresponding to a control statement) is compared with every row of program 2. Row I and j of the programs are compared using city block distance formula $|Ri1-Rj1|+|Ri2-Rj2|$.

For example first row of table 1 is compared with second row of table 2 by computing $|0-0| + |0-1| =1$ is entered in (1, 2) of distance matrix (table 3). From this table we can find similar control lines present in two programs. Presence of zero in a position corresponding to similar control statement indicates structural similarity of the control statements in the two programs. For example zero at (3, 1) in table 3 imply that the iterative statements 3 of program 1 and 1 of program 2 are probably similar. Whereas zero at (5, 3) is not comparable because the control statements of the programs are different (fifth control statement of program 1 is conditional and third control statement of program 2 is iterative). The zeros that contribute to similarity are highlighted.

Table 3. Distance matrix computed from table 1 and 2

Control lines	Loop (L)	Loop (L)	Loop (L)	Cond (C)	Loop (L)
Loop(L)	2	1	**0**	0	**0**
Loop (L)	2	1	**0**	0	**0**
Loop(L)	**0**	1	2	2	2
Loop(L)	1	**0**	1	1	1
Cond(C)	2	1	0	**0**	0
Loop(L)	2	1	**0**	0	**0**

Similarity between codes is found, using the formula

$$\begin{cases} s = n & \text{if} \quad r1=r2 \\ s = \dfrac{n}{|r1-r2|} & \text{otherwise} \quad \text{....... (1)} \end{cases}$$

where r1 and r2 are the number of control lines in two programs. From table 3 s = 9/1.

We conducted experiments using data set 1 of 5 distinct programs and 15 variants and similarity values are shown in table 4.

Table 4. Similarity table for data set 1 (s=n/|r1-r2|)

Programs	P1v1	P1v2	P1v3	P1v4	P2v1	P2v2	P2v 3	P3v1	P3v2	P3v3	P4v1	P4v2	P4v3	P5v1	P5v2
P1v 1	0.00	37.00	37.00	37.00	2.29	2.29	2.29	2.23	2.23	2.23	1.11	1.11	1.11	4.22	4.20
P1v2	37.00	0.00	37.00	37.00	2.29	2.29	2.29	2.23	2.23	2.23	1.11	1.11	1.11	4.22	4.20
P1v3	37.00	37.00	0.00	37.00	2.29	2.29	2.29	2.23	2.23	2.23	1.11	1.11	1.11	4.22	4.20
P1v4	37.00	37.00	37.00	0.00	2.29	2.29	2.29	2.23	2.23	2.23	1.11	1.11	1.11	4.22	4.20
P2v1	2.29	2.29	2.29	2.29	0.00	8.00	8.00	0.83	0.83	0.83	2.50	2.50	2.50	1.13	1.18
P2 2	2.29	2.29	2.29	2.29	8.00	0.00	8.00	0.83	0.83	0.83	2.50	2.50	2.50	1.13	1.18
P2v3	2.29	2.29	2.29	2.29	8.00	8.00	0.00	0.83	0.83	0.83	2.50	2.50	2.50	1.13	1.18
P3v1	2.23	2.23	2.23	2.23	0.83	0.83	0.83	0.00	199.00	199.00	0.61	0.61	0.61	10.92	13.83
P3v2	2.23	2.23	2.23	2.23	0.83	0.83	0.83	199.00	0.00	199.00	0.61	0.61	0.61	10.92	13.83
P3v3	2.23	2.23	2.23	2.23	0.83	0.83	0.83	199.00	199.00	0.00	0.61	0.61	0.61	10.92	13.83
P4v1	1.11	1.11	1.11	1.11	2.50	2.50	2.50	0.61	0.61	0.61	0.00	4.00	4.00	0.83	0.89
P4v2	1.11	1.11	1.11	1.11	2.50	2.50	2.50	0.61	0.61	0.61	4.00	0.00	4.00	0.83	0.89
P4v3	1.11	1.11	1.11	1.11	2.50	2.50	2.50	0.61	0.61	0.61	4.00	4.00	0.00	0.83	0.89
P5v1	4.22	4.22	4.22	4.22	1.13	1.13	1.13	10.92	10.92	10.92	0.83	0.83	0.83	0.00	161.00
P5v2	4.20	4.20	4.20	4.20	1.18	1.18	1.18	13.83	13.83	13.83	0.89	0.89	0.89	161.00	0.00

We may observe that in table 4 all programs show highest similarity only with its variants.

3.2 Approach 2: Computation of similarity using execution flow of control statements

In pre processing stage all functions are placed above the main function. Function Information
Table (FIT) and CST are generated in a single scan of the program.
Function Information Table (FIT): FIT gives starting and ending positions where a particular function begins and ends in CST. Here function calls are considered as a control lines. FIT of sort program 2 and 3 are shown in table 5a and 5b. CSTs of these programs are shown in table 6a and 6b.

Table 5a. Function Information Table (FIT) for sort program 2

Sl. No	Function name	Start position	End position
1	Sort	1	3
2	Print	4	4
3	main	5	8

Table 5b. Function Information Table (FIT) for sort program 3

Sl. No	Function name	Start position	End position
1	Sort	1	2
2	Print	3	3
3	Main	4	8

The line 1 (first control statement) of program 2 is function name 'sort' (beginning of function) is entered in FIT of table 5a (refer function name and start position). The control statements scanned from line 1 onwards are recorded sequentially in CST (table 6a) until end of the function. The end of the function namely line 3 is recorded in FIT. Thus in one scan FIT and CST are generated.

Execution Flow Control Structure Table (EFCST) is computed using CST and FIT by replacing the function calls by control lines of that particular function.

Table 6a. Control Structure Table (Order) for program 2 in figure 1

Sl. no	Control statement	Loop	Condition
1	Loop	1	1
2	Loop	0	1
3	Condition	0	0
4	Loop	0	0
5	Loop	0	0
6	Print	0	0
7	Sort	0	0
8	Print	0	0

Table 6b. Control Structure Table (Order) for program 3 in figure 1

Sl. no	Control statement	Loop	Condition
1	Loop	0	1
2	Condition	0	0
3	Loop	0	0
4	Loop	0	0
5	Print	0	0
6	Loop	0	0
7	Sort	0	0
8	Print	0	0

Execution Flow Control Structure Table (EFCST) of program 2 is given in table 7. Execution flow starts in 'main'. From FIT we see that flow starts at line 5 and ends at line 8. The entries in these lines are copied in EFCST. However if function call is present, FIT is referred as corresponding control lines of the function from the respective beginning and ending lines are copied to EFCST. The EFCST of programs 1, 2 and 3 in figure 1 are shown in table 7.

Table 7. EFCST of program 1, 2 and 3

Sl. no	Control statement	Loop	Condition
1	Loop	0	0
2	Loop	0	0
3	Loop	1	1
4	Loop	0	1
5	Condition	0	0
6	Loop	0	0

Difference matrix is computed using two EFCSTs as in section 3.1 and similarity value is computed using formula 1.

We conducted experiments on data set 1 and results are shown in below table. We conducted experiments on data set 1 and results are shown in table 8.

Table 8. EFCST and s=n/|r1-r2|

	P1v1	P1v2	P1v3	P1v4	P2v1	P2v2	P2v3	P3v1	P3v2	P3v3	P4v1	P4v2	P4v3	P5v1	P5v2
P1v1	0.00	36.00	36.00	36.00	2.14	2.14	2.14	2.14	2.14	2.14	1.00	1.00	1.00	3.55	3.30
P1v2	36.00	0.00	36.00	36.00	2.14	2.14	2.14	2.14	2.14	2.14	1.00	1.00	1.00	3.55	3.30
P1v3	36.00	36.00	0.00	36.00	2.14	2.14	2.14	2.14	2.14	2.14	1.00	1.00	1.00	3.55	3.30
P1v4	36.00	36.00	36.00	0.00	2.14	2.14	2.14	2.14	2.14	2.14	1.00	1.00	1.00	3.55	3.30
P2v1	2.14	2.14	2.14	2.14	0.00	7.00	7.00	0.76	0.76	0.76	2.00	2.00	2.00	1.00	0.88
P2v2	2.14	2.14	2.14	2.14	7.00	0.00	7.00	0.76	0.76	0.76	2.00	2.00	2.00	1.00	0.88
P2v3	2.14	2.14	2.14	2.14	7.00	7.00	0.00	0.76	0.76	0.76	2.00	2.00	2.00	1.00	0.88
P3v1	2.14	2.14	2.14	2.14	0.76	0.76	0.76	0.00	196.00	196.00	0.58	0.58	0.58	13.91	9.83
P3v2	2.14	2.14	2.14	2.14	0.76	0.76	0.76	196.00	0.00	196.00	0.58	0.58	0.58	13.91	9.83
P3v3	2.14	2.14	2.14	2.14	0.76	0.76	0.76	196.00	196.00	0.00	0.58	0.58	0.58	13.91	9.83
P4v1	1.00	1.00	1.00	1.00	2.00	2.00	2.00	0.58	0.58	0.58	0.00	3.00	3.00	0.75	0.63
P4v2	1.00	1.00	1.00	1.00	2.00	2.00	2.00	0.58	0.58	0.58	3.00	0.00	3.00	0.75	0.63
P4v3	1.00	1.00	1.00	1.00	2.00	2.00	2.00	0.58	0.58	0.58	3.00	3.00	0.00	0.75	0.63

P5v1	3.55	3.55	3.55	3.55	1.00	1.00	1.00	13.91	13.91	13.91	0.75	0.75	0.75	0.00	125.00
P5v2	3.30	3.30	3.30	3.30	0.88	0.88	0.88	9.83	9.83	9.83	0.63	0.63	0.63	125.00	0.00

Here also all programs show high similarity only with versions of the same program.

3.3 Similarity computation using CSTs, EFCSTs and Control Metric Table (CMT)

Control Metric Table (CMT): We compute control metric table which contains information about total number of iterative and conditional statements present in the program [11]. Table 9 shows CMT of data set 1 used for our experiment.

Table 9. Control Metric Table for data set 1 (CMT)

Sl. No	Programs	1		2		3		4	
		L	**C**	**L**	**C**	**L**	**C**	**L**	**C**
1	Beam search	10	2	10	2	10	2	10	2
2	Bubble sort	4	1	4	1	4	1	-	-
3	Min Max	15	19	15	19	15	19	-	-
4	Linear search	2	1	2	1	-	-	-	-
5	Queue	3	18	3	18	3	18	-	-

Computation of similarity value (s): Here similarity computation is based on CMT as well as CST/EFCST. First we generate CMT and CST for each program. Difference matrix (D) is computed from the respective CSTs as explained in earlier sub sections 3.1 and 3.2.

We compute similarity between programs only if programs are comparable in terms of number of loops and conditional statements. While duplicates are created it is unlikely to expect more than 20 % variation in number of control statements. Hence a threshold of 20 % variations in these numbers is fixed for computation of similarity. Suppose program 1 has x loops and y conditional statements. Program 2 is comparable with program 1 if the number loops and conditional statements are in the range [x – 20 % (x), x + 20 % (x)] and [y – 20 % (y), y + 20 % (y)]. Table 10 show computed similarity values with this additional consideration of CMT.

Table 10a. CST, CMT and s=n/|r1-r2|

	P1v1	**P1v2**	**P1v3**	**P1v4**	**P2v1**	**P2v2**	**P2v3**	**P3v1**	**P3v2**	**P3v3**	**P4v1**	**P4v2**	**P4v3**	**P5v1**	**P5v2**
P1v1	0	37	37	37	0	0	0	0	0	0	0	0	0	0	0
P1v2	37	0	37	37	0	0	0	0	0	0	0	0	0	0	0
P1v3	37	37	0	37	0	0	0	0	0	0	0	0	0	0	0
P1v4	37	37	37	0	0	0	0	0	0	0	0	0	0	0	0
P2v1	0	0	0	0	0	8	8	0	0	0	0	0	0	0	0
P2v2	0	0	0	0	8	0	8	0	0	0	0	0	0	0	0
P2v3	0	0	0	0	8	8	0	0	0	0	0	0	0	0	0
P3v1	0	0	0	0	0	0	0	0	199	199	0	0	0	0	0
P3v2	0	0	0	0	0	0	0	199	0	199	0	0	0	0	0
P3v3	0	0	0	0	0	0	0	199	199	0	0	0	0	0	0
P4v1	0	0	0	0	0	0	0	0	0	0	0	4	4	0	0
P4v2	0	0	0	0	0	0	0	0	0	0	4	0	4	0	0
P4v3	0	0	0	0	0	0	0	0	0	0	4	4	0	0	0
P5v1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	161
P5v2	0	0	0	0	0	0	0	0	0	0	0	0	0	161	0

Table 10b. EFCST, CMT and s= n/|r1-r2|

	P1v1	**P1v2**	**P1v3**	**P1v4**	**P2v1**	**P2v2**	**P2v3**	**P3v1**	**P3v2**	**P3v3**	**P4v1**	**P4v2**	**P4v3**	**P5v1**	**P5v2**
P1v1	0	36	36	36	0	0	0	0	0	0	0	0	0	0	0
P1v2	36	0	36	36	0	0	0	0	0	0	0	0	0	0	0
P1v3	36	36	0	36	0	0	0	0	0	0	0	0	0	0	0
P1v4	36	36	36	0	0	0	0	0	0	0	0	0	0	0	0

P2v1	0	0	0	0	0	7	7	0	0	0	0	0	0	0	0
P2v2	0	0	0	0	7	0	7	0	0	0	0	0	0	0	0
P2v3	0	0	0	0	7	7	0	0	0	0	0	0	0	0	0
P3v1	0	0	0	0	0	0	0	0	196	196	0	0	0	0	0
P3v2	0	0	0	0	0	0	0	196	0	196	0	0	0	0	0
P3v3	0	0	0	0	0	0	0	196	196	0	0	0	0	0	0
P4v1	0	0	0	0	0	0	0	0	0	0	0	3	3	0	0
P4v2	0	0	0	0	0	0	0	0	0	0	3	0	3	0	0
P4v3	0	0	0	0	0	0	0	0	0	0	3	3	0	0	0
P5v1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	125
P5v2	0	0	0	0	0	0	0	0	0	0	0	0	0	125	0

In the above tables similarity is seen only with versions of the same program. All others are 0's.

3.4 Experimental Results

Five programs, 15 versions data set described in earlier sections is created in our lab and the experimental results with two approaches have been discussed in detail in sections 3.1 to 3.3.

For thorough testing of the proposed approaches we downloaded programs from 'sourcefoge.net' (www.sourceforge.net) and 'f1sourcecode' (www.f1sourcecode.com) and created many versions by changing loop statements, reordering control lines and also by refactoring. These are added to the sample data set in the earlier sections. Thus we have created 26 distinct programs and 100 versions data set. To find whether only versions of the same programs, show higher similarity when compared to similarities with other programs, we have done clustering of similarity values using k-means clustering algorithm with k=2. The clustering is done on set of similarity value corresponding to one version of a

program (available in a column). The error in duplicate detection of a program 'j' is found as ratio of number of misclassification and total number of programs (inclusive of versions). Total misclassification in program 'j' includes number of false positives and true negatives. When a version of program 'j' is clustered with any other program it is true negative, where as when a version of program 'i' is clustered with program 'j' it is false positive.

Average error is computed total detection errors in each program by number of distinct programs. Table 11 shows the average error with two approaches with and without CMT for the sample data sets. Also shown in the table the similarity measurements using the formula s=n/D, where n is similar number of control lines and 'D' maximum dissimilarity [11].

Table 11. Error table for sample data sets.

Approaches	Data structure used	Data set1		Data set 2	
		S= n /D	S= n / |r1-r2|	S= n / D	S=n / |r1-r2|
Approach 1	Only CST	0.1465	0.0375	0.5794	0.1038
	CST and CMT	0	0	0.00923	0.00577
Approach 2	Only EFCSTs	0.04	0.0375	0.0866	0.009615
	EFCST and CMT	0	0	0.009615	0.00808

3.5 Time Complexity

Suppose two programs have n_1 and n_2 source lines and L_1 and L_2 control statements. Note that number of control statements in a program will be far less than number of source lines ($L << n$). Table 12 shows the detail of major steps in the computation of similarity and the corresponding complexities.

Table 12. Time complexity table

Steps	Complexity
Preprocessing	$\theta(n_1) + \theta(n_2)$
CST / EFCST	$\theta(n_1) + \theta(n_2)$
Difference matrix	$\theta(L_1 \times L_2)$
Similarity computation	$O(L_1 \times L_2)$

Hence total time complexity is maximum ($\theta(n)$ and $O(L^2)$) which is a polynomial time complexity.

3.6 Performance Evaluation

The experiments are done with three available tools **Duplo** (uses string matching technique), **PMD** (uses tokens to compare) and **CloneDR** (AST based) and the results obtained on data set 1 is shown in table 13.

PMD tool shows similarity with user defined function call and inbuilt function. Control lines for and while, from figure 1 are not shown as similar. CloneDR is sometimes sensitive to change in the type of loop statement.

We divided data set 2 which is used in section 3.4 into two data sets. First data set has 15 distinct programs and 50 variants. This data set has variation in sequence of control statements (independent control lines only) in versions of the same program. Second data set has 11 distinct programs and 50 variants. In this data set contents of control lines are replaced by function calls (refer fig 1).

Experiments are conducted on two data sets using two approaches. Tables 14a and 14b show performance analysis for proposed methods.

Table 13. Performance analysis table

Sl. no	Method	Error		Remarks
1	Duplo	1.8666		All versions of beam search show some similarity with all versions of minmax and bubble sort programs are not shown as similar programs.
2	PMD	1.6		All versions of beam search show some similarity with all versions of minmax and queue programs are not shown as similar programs.
3	Clone DR	1.8666		All versions of beam search show some similarity with all versions of minmax and queue programs are not shown as similar programs.
4	Proposed Approaches	Only CST	0.14658	Linear search and beam search programs show similarity with versions of other programs
		Only EFCST	0.04	Linear search program shows similarity with bubble sort programs
		CST & CMT	0	Similarity exists with its versions only
		EFCST & CMT	0	

Table 14a. Performance analysis table (without considering CMT)Without CMT

Data structure and similarity measure used	Data set1	Data set 2	Data set 3
CST & s=n/d	0.14658	0.34	0.292727
CST & s=n/|r1-r2|	0.0375	0.0866	0.092727
EFCST & s=n/|r1-r2|	0.0375	0.0373	0.049
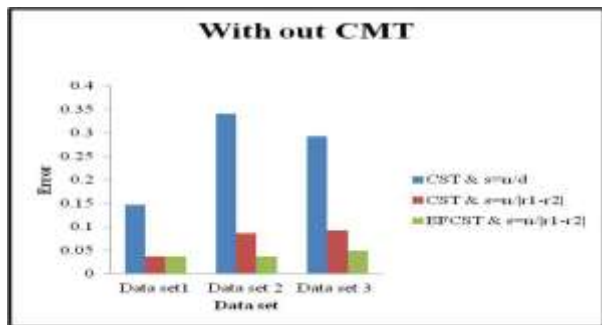
Fig 4: Error graph for proposed approaches without considering CMT

Table 14b. Performance analysis table (considering CMT)

Data structure and similarity measure used	Data set1	Data set 2	Data set 3
CST & s=n/d	0	0.0133	0.0436
CST & s=n/|r1-r2|	0	0.00933	0.02
EFCST & s=n/|r1-r2|	0	0	0
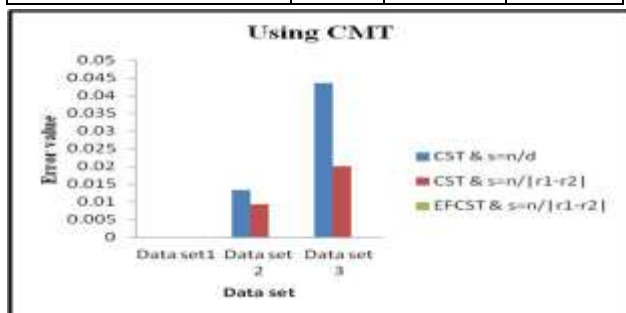
Fig 5: Error graph for proposed approaches without considering CMT

4. CONCLUSION AND FUTURE WORK

We have proposed two approaches Control Structure Table (CST) and Execution Flow Control Structure Table (EFCST) to detect duplicate code detection. We also suggested Control Metric Table (CMT) before computation of similarity measure. Performance with the addition of CMT has shown tremendous improvements.

The time complexity is max ($\theta(n)$ and $O(L^2)$) where 'n' is total number of source lines and 'L' is total number of control statements in the program. Time complexity is far less when compared to methods based on AST and PDG. The method also identifies all four types of clones.

The proposed algorithms do not take into consideration of statements inside control structures. The current similarity measure can be corrected to consider the statements together with operators and operands. Perhaps errors that are observed currently may decrease significantly.

5. REFERENCES

[1] Baker S., "A Program for Identifying Duplicated Code, "Computing Science and Statistics, vol. 24, 1992.

[2] Johnson J H., "Substring matching for clone detection and change tracking," in Proceedings of the International Conference on software Maintenance, 1994.

[3] Ducasse, S., Rieger M., and Demeyer S., "A Language Independent Approach for Detecting Duplicated Code." In Proceedings; IEEE International Conference on Software Maintenanace, 1999.

[4] Zhang Q., et . al., "Efficient Partial-Duplicate Based on Sequence Matching," 2010.

[5] Sadowski C., and Levin G., "SimHash: Hash-Based Similarity Detection," 2007.

[6]Jiang L., and Glondu S., "Deckard: Scalable and Accurate Tree-Based Detection of Code Clones".

[7] Baxter I D., Yahin I., Moura L., Anna M S., and Bier L., "Clone Detection Using Abstract Syntax Trees," in proceedings of ICSM. IEEE, 1998.

[8] krinke J., "Identifying Similar Code with Program Dependency Graphs," Proc. Eighth Working Conference ., Reverse Engineering., 2001.

[9] Vidya K and Thirukumar K, "Identifying Functional Clones between Java Directory using Metric Based Systems" International journal of Computer Communication and Information System (IJCCI)-Vol3, ISSN:2277-128x August 2013.

[10] Mayrand J, Leblanc C and Ettore Merlo M. "Experiment on the Automatic Detection of Function Clones in a Software System Using Metrics", proc of ICSM conference 1996.

[11]Sudhamani M and Rangarajan L, Structural Similarity Detection using Structure of Control Statements, proc. of International Conference on Information and Communication Technology, vol 46 (2015) 892-899.

[12] kodhai E, Perumal A, and Kanmani S, "Clone Detection using Textual and metric Analysis to figure out all Types of Clones" International journal of Computer Communication and Information System (IJCCI)- Vol2. No1.ISSN:0976-1349 July-Dec 2010.

[13] www.research.cs.queensu.ca.

[14] Bellon S., Koschke R., Antoniol G., Krinke J., and Merlo E., "Comparison and evaluation of clone detection tools," IEEE Transactions on.Software Engineering, September 2007.

[15]. Roy C K and Cordy J R. "A survey on software clone detection research". Tech. rep., 2007. TR 2007-541 School of Computing Queen's University at Kingston Ontario, Canada.

[16] Roy C K, Cordy J R and Koschke R, "Comparison and Evaluation of Code Clone Detection Techniques and Tools: A Qualitative Approach", Science of Computr Programming, 74(2009) 470-495, 2009.

[17] Kostas Kontogiannis. "Evaluation Experiments on the Detection of Programming Patterns using Software Metrics". In Proceedings of the 3rd Working Conference on Reverse Engineering (WCRE'97), pp. 44-54, Amsterdam, the Netherlands, October 1997.

Software Architecture Evaluation of Unmanned Aerial Vehicles Fuzzy Based Controllers – A Survey

Elsiddig ahmed Elsiddig
College of Graduate Studies
Sudan University of Science and Technology
Khartoum, Sudan

Hany H Ammar
Lane Department of Computer Science and Electrical Engineering
College of Engineering and Mineral Resources
West Virginia University, Morgantown, WV, USA

Abstract: In this survey paper we discuss the recent techniques for software architecture evaluation methods for Unmanned Aerial Vehicle (UAV) systems that use fuzzy control methodology. We discuss the current methodologies and evaluation approaches, identify their limitations, and discuss the open research issues. These issues include methods used to evaluate the level of risk, communications latency, availability, sensor performance, automation, and human interaction.

Keywords: evaluation methods of UAVs, Fuzzy Based Controllers, autonomy evaluation method, full autonomy, performance.

1. INTRODUCTION

Researchers around the world have carried much research into the evaluation of unmanned systems, each having their own clear application objectives. Our review of the existing software architecture evaluation methods of the autonomy of unmanned systems has found that these methods include the level evaluation method, double-axis method, three-axis method, look-up table method and formulation method.

There are many different researches work relate to the software architecture evaluation of autonomous navigation control of UAVs; and it content different techniques to develop UAVs control system, **include :**

      a) Fuzzy control technique.
      b) Adaptive control technique.
      c) Neural networks technique.
      d) Genetic algorithms technique.

Then in this survey we focus on the papers that using fuzzy control technique to control all system components.

Fuzzy logic (FL) will be used to tune the PD or PID control gains according to the readings of the sensors, which monitors the condition of the structure and the engine. Since the development of such a hybrid system proves to be time consuming and involves generating many algorithms and models, we only focus on one strand of the hybrid scheme, which is the fuzzy logic control in this research survey. This artificial intelligence system is quite favored for automatic control because it avoids complex non-linear equations and can utilize the best pilot expertise available. Fuzzy logic that is based on the mathematical theory of fuzzy sets circumvents complex differential equations by offering a collection of if-then rules that operates as a linear function even though the function itself is not known. and this logic does not operate on binary output such as true or false, up or down, left or right, but rather facilitates for the entire intermediate spectrum of outputs to be included. In this research survey, previous fuzzy logic controllers (FLC) are reviewed. [10]

2. SURVEY OF THE CURRENT SOFTWARE ARCHITECURE EVALUATION METHODS

Several classical software architecture methods have been widely applied and cited, including *Sheridan's*LOA[22, 25], the US military's Autonomous Control Level (ACL)[22] , and Huang's ALFUS[22]. Researchers have extended the above software architecture evaluation methods to various degrees.

For example, using Yang and Zhang [22] put forward the assessment of the autonomy level of unmanned systems by considering four aspects the degree of change in the environment, the degree of task completion, the degree of state stability of the system, and the degree of human computer interaction and then introduced fuzzy theory to this method by quantifying the autonomy level of unmanned systems with a fuzzy decision. Nevertheless the UAV components are not controlled by fuzzy control technique.

From analysis of the mechanism of human intelligence and with in-depth study of the autonomy control-level classification of the UAV, Chen et al.[22] proposed measures of nine-level that evaluate the autonomy control of an unmanned system, and then enriched the content of the autonomy control evaluation level with an autonomy function, autonomy types, intelligence attributes, information sharing capabilities and other aspects. HOWEVER there is no using of fuzzy theory.

Suresh and Ghose [23] carried out measures on 11 levels of autonomy control from single system autonomy to group autonomy, and they provided detailed explanations of each level through published extensive literature on autonomy. They divided each level into several sub layers, and they believed that communication and information would play a key role in realizing the autonomy of an UAV. However there is no using of fuzzy theory.

Many of papers are concerned with the flight of UAVs. proposes fuzzy logic based autonomous flight and landing

system controllers. All of them using the same three fuzzy logic modules and these modules are developed under the main navigation control system and three more for the autonomous landing control system to control of the altitude, the speed, and the position against the runway, through which the global position (latitude-longitude) of the air vehicle is controlled. A SID (Standard Instrument Departure) and TACAN (Tactical Air Navigation) approach is used and the performance of the fuzzy-based controllers is evaluated with time based diagrams under MATLAB's standard configuration and the (Aerosim) Aeronautical Simulation Block Set which provides a complete set of tools.

The Aerosonde UAV model is **always** used in the simulations in order to demonstrate the performance and the potential of the controllers. Additionally, some visual tools are deployed in order to get visual outputs that help the designer in the evaluation of the controllers. [4][5][8]

SeferKurnaz ·Omer Cetin · OkyayKaynak describes other design of compact, and inexpensive fuzzy logic controllers and fuzzy inference systems which estimates the attitude of (UAV), and this attitude refers to parameters of UAV such as longitude latitude, and altitude and angles of rotation known as pitch and roll.

Fuzzy Logic is used to design the Fuzzy Logic Controllers and Fuzzy Inference Systems. Visual simulation tool and Aerosim (Aeronautical Simulation Set) Flight gear interface are used for Simulation purpose.[5]However one UAV mission model used in the simulations

*L. Doitsidis, K. P. Valavanis*they has been developed two module fuzzy logic controller that also includes a separate error calculating box is derived for autonomous navigation and control of small manned – unmanned aerial vehicles demonstrating ability to fly through specified way-points in a 3-D environment repeatedly, perform trajectory tracking, and, duplicate another vehicle's trajectory. A *MATLAB* standard configuration environment and the Aerosim Aeronautical Simulation Block Set are utilized for simulation studies, presented through a visualization interface; results illustrate controller performance and potential. However this design is currently based on human pilot experience and not on flight performance observations. [7].

Kimberly Bickraj discusses the integration of health monitoring and flight control systems for small (UAVs). After briefly reviewing previous fuzzy logic controllers (FLC) of air vehicles, there is very low cost integration method is proposed. The proposed fuzzy logic (FL) selects the best gain values for the operation of PD or PID controllers of the autonomous flight system according to the health of the components. any gain adjustments help the UAV to execute maneuvers in a more conservative manner when the system have structural or proportion system problems. However Once this is successfully established, we will be one-step closer to achieving a completely self- sufficient UAV. [9]

James K. Kuchar civil airspace requires new methods of ensuring collision avoidance. Concerns over command and control latency, vehicle performance, reliability of autonomous functions, and interoperability of sense-and-avoid systems with the Traffic Alert and Collision Avoidance System (TCAS) and Air Traffic Control must be resolved. his paper describes the safety evaluation process that the international community has deemed necessary to certify such systems. The process focuses on a statistically-valid estimate of collision avoidance performance developed through a combination of airspace encounter modeling, fast time simulation of the collision avoidance system across millions of encounter scenarios, and system failure and event sensitivity analysis.

However, all aircraft are required to be transponder -equipped. Extensive flight testing is required to support modeling communications latency and availability, sensor performance, automation, human interaction with CAS advisories, and flight characteristics.[1]

Meng-Lun fixed-wing UAV based spatial information acquisition platform is developed and evaluated .the evaluation Approach used is comparing DG results with checking points with precisely known coordinates. To evaluated direct Georeference(DG) accuracy, Interior Orientation Parameters (IOP) and Exterior Orientation Parameters (EOP). However the Extended Kalman Filter(EKF) trajectory is not smooth [2].

*DemozGebre*develop a framework for the design of CONOPs, which take these SUAS limitations into account. The method outlined shows, in part, how these vehicle/infrastructure collision risks can be estimated or conservatively bounded. his evaluation Approach used is resorting and multi-sensor on board to evaluate Risk, vehicle states, traffic management parameters, and sensor performance. However the performance specifications provided by manufactures does not provide sufficient information to allow precisely quantifying or bounding the collision risk.[3]

Roopashree.S, Shubha Bhat. they estimates the attitude of UAVs. they used Kalman filter fuzzy approach to generate fuzzy rules. (Kalman Filtering is a form of optimal estimation characterized by recursive evaluation) and they have internal model of dynamics of system being estimated. However The attitude of UAV may oscillate because the controller design is based on human pilot experience.[6]

3. OPEN ISSUES

After having studied the literature on evaluating autonomous control of UAVs we can say that :

a. There was no systematic study that could provide a full autonomy evaluation method for fuzzy logic based control system of UAVs.

b. There are still many problems with the adaptability of existing evaluation methods, which need to be improved for *general utilization*.

c. The autonomy evaluation method of the UAV must consider the *diversity*, *multidimensionality*, *hierarchy*, and primary and secondary natures of the application target and the system itself, so that it can

avoid *ambiguity* in the formulation and description
of the autonomy of autonomous UAV.

**Table 1. Examples of Approaches for Software
Architecture Evaluation of UAVs:**

	Author Name/Year	Objective	evaluation Approach used	Elements or attributes that evaluated	Limitation
1	SeferKurnaz ·Omer Cetin · OkyayKaynak(2009)	provides autonomy to the UAV in all phases of a typical UAV's mission except take off and land on.	MATLAB's standard configuration and the Aerosim Aeronautical Simulation Block Set	performance of the fuzzy based controllers components	there exist some oscillations and errors when wind effects are added to the simulation environment one UAV mission model used in the simulations
2	Omer Cetin · SeferKurnaz · OkyayKaynak (2011)	Design of Autonomous Landing System for Unmanned Aerial Vehicles	MATLAB's standard configuration and the Aerosim Aeronautical Simulation Block	performance of the fuzzy based controllers components	there exist some oscillations and errors when wind effects are added to the simulation environment.. one UAV mission model used in the simulations
3	SeferKurnaz , Omer Çetin (2010)	Autonomous Navigation and Landing Tasks for Fixed Wing Small (UAV)	MATLAB's standard configuration and the Aerosim Aeronautical Simulation Block	performance of the fuzzy based controllers components	low tuning of membership functions. Autonomous takeoff not tried. one UAV mission model used in the simulations
4	L. Doitsidis, K. P. Valavanis (2004)	ability to fly through specified waypoints in a 3-D environment repeatedly	Fuzzy Logic Based	controller performance	Design is currently based on human pilot experience and not on flight performance observations.
5	KimberlyBickraj,Thierry Pamphile (2006)	selects the best gain values for the operation of PD or PID controllers of the autonomous flight system according to the health of the components.	fuzzy logic based neuro-fuzzy inference system (ANFIS).	performance of the fuzzy based controllers components	Once this is successfully established, we will be one-step closer to achieving a completely self- sufficient UAV.
6	James K. Kuchar	Safety analysis methodology for (UAV) collision avoidance systems	statistically-valid estimate	Risk ratio, communications latency, availability, sensor performance, automation, human interaction .	1- UAV concepts need to go through a similar process involving detailed airspace encounter modeling, dynamic simulation of collision avoidance system performance, and system failure and event sensitivity studies. 2 -Flight tests and human-in-the-loop simulation studies are also required to develop models to describe how UAV pilots (or an autonomous system) would respond to collision avoidance

					system advisories.
7	Meng-Lun (2012)	fixed-wing UAV based spatial information acquisition platform is developed and evaluated.	comparing DG results with checking points with precisely known coordinates	DG accuracy, Interior Orientation Parameters (IOP) Exterior Orientation Parameters (EOP)	EKF trajectory is not smooth
8	DemozGebre-EgziabherZhiqiang Xing 2011	develop a framework for the design of CONOPs, which take these SUAS limitations into account. The method outlined shows, in part, how these vehicle/infrastructure collision risks can be estimated or conservatively bounded.	Resorting. Multi- sensor.	Risk. vehicle states . other traffic management parameters. sensor performance.	The performance specifications provided by manufactures does not provide sufficient information to allow precisely quantifying or bounding the collision risk.
9	Roopashree.S, Shubha Bhat (2012)	estimates the attitude of (UAV)	Kalman filter fuzzy approach is used to generate fuzzy rules. Kalman Filtering is a form of optimal estimation characterized by recursive evaluation	internal model of dynamics of system being estimated.	The attitude of UAV may oscillate because the controller design is based on human pilot experience.

4. CONCLUSION

This survey paper concern with software architecture evaluation of unmanned aerial vehicles fuzzy based controllers.

In the literature many approaches to architecture evaluation can be identified, i.e., simulation, scenarios, mathematical modeling, and experience based reasoning. Here we present some evaluation approach and what are the elements or attributes that evaluated and what is there limitation. Then we find an interesting open issue about the evaluation on autonomous degree and performance of system controller's components of UAVs.

From our survey there are still many problems with the adaptability of existing evaluation methods, which need to be improved for general utilization. The autonomy evaluation

method of the UAV must consider the *diversity*, *multidimensionality*, *hierarchy*, and primary and secondary natures of the application target and the system itself, so that it can avoid *ambiguity* in the formulation and description of the autonomy of autonomous UAV.

5. REFERENCES

[1] James K. Kuchar, MIT Lincoln Laboratory" SAFETY ANALYSIS METHODOLOGY FOR UNMANNED

AERIALVEHICLE (UAV) COLLISION AVOIDANCE SYSTEMS".

[2] Meng-Lun Tsai a, *, Kai-Wei Chiang a, Yi-Hsing Tseng a, Jiann-You Rau a, Yun-Wen Huang a, Cheng-Fang Lo b "THE PERFORMANCE ANALYSIS OF A UAV BASED MOBILE MAPPING SYSTEM " Dept. of Geomatics, National Cheng Kung University, Taiwanb GeoSat Informatics Technology Corporation, Taiwan

[3] DemozGebre-EgziabherZhiqiang Xing March 2011 " Analysis of Unmanned Aerial Vehicles Concept of Operations in ITS Applications" Department of Aerospace Engineering & Mechanics University of Minnesota

[4]. SeferKurnaz ·Omer Cetin · OkyayKaynak." Fuzzy Logic Based Approach to Design of Flight Control and Navigation Tasks for Autonomous Unmanned Aerial Vehicles"Received:

15 March 2008 / Accepted: 30 June 2008 / Published online: 1 October 2008 © Springer Science + Business Media B.V. 2008

[5]. Omer Cetin 1, SeferKurnaz 2, Okyay Kaynak1,3 "Fuzzy Logic Based Approach to Design of Autonomous Landing System for Unmanned Aerial Vehicles " 1. Turkish Air Force Academy, ASTIN o.cetin@hho.edu.tr, Yesilyurt, Istanbul, 34807 Turkey   2. Turkish Air Force Academy, ASTIN Yesilyurt, Istanbul, 34807 Turkey 3. Department of Electrical and Electronic Engineering, Bogazici University, o.kaynak_at_ieee.org, Bebek, 80815 Istanbul / Turkey

[6]Roopashree.S #1, Shubha Bhat#3:Deepika K.M #2 "Fuzzy Kalman Approach to Control the Navigation of Unmanned Aerial Vehicles" 1, 3 Department of Computer Science and Engineering, Dayananda Sagar College of Engineering, Affiliated to VTU, Bangalore-560078, India.2 Department of Information Science and Engineering, NitteMeenakshiInstiute of Technology, Affiliated to VTU, Bangalore-560064, India.

[7] L. Doitsidis, K. P. Valavanis 1 N. C. Tsourveloudis, M. Kontitsis2 "A Framework for Fuzzy Logic Based UAV Navigation and Control 1- Dept. of CSEE, CRASAR University of South Florida Tampa, FL 33620 2- Dept. of Production Engineering and Management Technical University of Crete Chania 73100, Crete, Greece

[8]. SeferKurnaz, Omer Çetin "Autonomous Navigation and Landing Tasks for Fixed Wing Small Unmanned Aerial Vehicles" Turkish Air Force Academy, ASTIN Yesilyurt, Istanbul, 34807 Turkey

[9]. Kimberly Bickraj 1,Aylin Yenilmez2, Ming Li, Ibrahim N3 "Fuzzy Logic Based Integrated Controller for Unmanned Aerial Vehicles"1. Thierry Pamphile Florida International University Department of Mechanical and Materials Engineering .2. Istanbul Technical University Faculty of Mechanical EngineeringGumussuyu, Istanbul 34439 Turkey 3. Tansel Florida International University Department of Mechanical and Materials Engineering

[10]. SWETHA PANDHITI"Blackboard Architecture for An Unmanned Aerial Vehicle Controller Using Fuzzy Inference Systems." B.Tech, Jawaharlal Nehru Technological University Hyderabad, Andhra Pradesh, India.

[11]. Dufrene, W.R., Jr. "Application of artificial intelligence techniques in uninhabited aerial vehicle flight. " In: The 22nd Digital Avionics Systems Conference vol. 2, pp. 8.C.3–8.1-6 (2003)

[12]. Li, Y., Sundararajan, N., Sratchandran, P. "Neuro-controller design for nonlinear fighter aircraft maneuver using fully tuned RBF networks" . Automatica 37, 1293–1301 (2001)

[13]. Borrelli, F., Keviczky, T., Balas, G.J "Collision-free UAV formation flight using decentralized optimization and invariant sets". In: 43rd IEEE Conference on Decision and Control vol. 1, pp. 1099–1104 (2004)

[14].Marin, J.A., Radtke, R., Innis, D., Barr, D.R., Schultz,A.C" Using a genetic algorithm to develop rules to guide unmanned aerial vehicles" In: Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, vol. 1, pp. 1055–1060. (1999)

[15]. Ren, W., Beard, R.W." CLF-based tracking control for UAV kinematic models with saturationconstraints". In: Proceedings of the 42nd IEEE Conference on Decision and Control, vol. 4, pp.3924–3929 (2003)

[16]. Dathbun, D., Kragelund, S., Pongpunwattana, A., CapozziB.: "An evolution based path planning algorithm for autonomous motion of a UAV through uncertain environments" .In: Proceedings of the 21st Digital Avionics Systems Conference vol. 2, pp. 8D2-1–8D2-12 (2002)

[18].Global Robotic Observation System, Definition OfAerosonde UAV Specifications: http://www.aerosonde.com/aircraft/

[19].Unmanned Dynamics, Aerosim Aeronautical Simulation Block Set Version 1.2 User's Guide: http://www.u-dynamics.com/aerosim/default.htm

[20]. FlightGear Open-source Flight Simulator. www.flightgear.org

[21]. Qiau, W., Muzimoto, M.: "PID type fuzzy controller and parameter adaptive method". Fuzzy Sets Syst. 78, 23–35 (1995),

[22] Huang H M. In: Messina E, Madhavan R, eds"Autonomy Levels for Unmanned Systems (ALFUS) framework: Safety and application issues.." Proceedings of the Performance Metrics for Intelligent Systems Workshop, 2007 Aug 28-30, Washington DC USA, ACM New York NY USA Press, 2007. 48–53

[23] Suresh M, Ghose D"Role of information and communication in redefining unmanned aerial vehicle autonomous control levels.." P I MechEng G-J Aer, 2010, 224: 171–197

[24] , Y. Zhao, Y. Yao, "Autopilot Design for Aerial Vehicles with Aerodynamic Surfaces and Lateral Jets Using Explicit Hybrid" MPC 1600 hrs AIAA-2015-0864 B. YangHarbin Institute of Technology, Harbin, China 101

[25] Defense Science Board "TASK FORCE REPORT: The Role of Autonomy in DoD Systems," Department of Defense, July 2012.

6. AUTHORS BIOGRAPHIES

Elsiddig ahmed Elsiddig
College of Graduate Studies
Sudan University of Science and Technology,
Khartoum, Sudan a PhD candidate in computer science since 2013.

Hany H. Ammar BSEE, BSPhysics, MSEE, and PhD EE, is a Professor of Computer Engineering in the Lane Computer Science and Electrical Engineering department at West Virginia University. He has published over 170 articles in prestigious international journals and conference proceedings. He is currently the Editor in Chief of the Communications of the Arab Computer Society On-Line Magazine. He is serving and has served as the Lead Principal Investigator in projects funded by the Qatar National Research Fund under the National Priorities Research Program. In 2010 he was awarded a Fulbright Specialist Scholar Award in Information Technology funded by the US State Department - Bureau of Education and Cultural Affairs. He has been the Principal Investigator on a number of research projects on Software Risk Assessment and Software Architecture Metrics funded by NASA and NSF, and projects on Automated Identification Systems funded by NIJ and NSF. He has been teaching in the areas of Software Engineering and Computer Architecture since 1987. In 2004, he co-authored a book entitled Pattern-Oriented Analysis and Design: Composing Patterns to Design Software Systems, Addison-Wesley. In 2006, he co-authored a book entitled Software Engineering: Technical, Organizational and Economic Aspects, an Arabic Textbook. Dr. Ammar is a member of the IEEE and the ACM professional organizations. He has served as Chairman and member of Steering Committees and Program Committees of several International Conferences, Workshops, and Symposia.

Integration of Bayesian Theory and Association Rule Mining in Predicting User's Browsing Activities – Survey Paper

Geoffrey Gitonga
Department of Computing
School of Computing and
Information Technology
Jomo Kenyatta University of
Agriculture & Technology
Nairobi, Kenya

Wilson Cheruiyot
Department of Computing
School of Computing and
Information Technology
Jomo Kenyatta University of
Agriculture & Technology,
Nairobi, Kenya

Waweru Mwangi
Department of Computing
School of Computing and
Information Technology
Jomo Kenyatta University of
Agriculture & Technology,
Nairobi, Kenya

Abstract: Bayesian theory and association rule mining methods are artificial intelligence techniques that have been used in various computing fields, especially in machine learning. Internet has been considered as an easy ground for vices like radicalization because of its diverse nature and ease of information access. These vices could be managed using recommender systems methods which are used to deliver users' preference data based on their previous interests and in relation with the community around the user. The recommender systems are divided into two broad categories, i.e. collaborative systems which considers users which share the same preferences as the user in question and content-based recommender systems tends to recommend websites similar to those already liked by the user. Recent research and information from security organs indicate that, online radicalization has been growing at an alarming rate. The paper reviews in depth what has been carried out in recommender systems and looks at how these methods could be combined to from a strong system to monitor and manage online menace as a result of radicalization. The relationship between different websites and the trend from continuous access of these websites forms the basis for probabilistic reasoning in understanding the users' behavior. Association rule mining method has been widely used in recommender systems in profiling and generating users' preferences. To add probabilistic reasoning considering internet magnitude and more so in social media, Bayesian theory is incorporated. Combination of this two techniques provides better analysis of the results thereby adding reliability and knowledge to the results.

1. INTRODUCTION

The use of internet has been growing tremendously over the years accelerated by the ease in internet access. This has led to growing dependent on the internet for chores like business, work and other social responsibilities like sharing of information, files and opinions using applications like Facebook, Twitter and Friendster among others. As a result of this enormous development of the internet and web based applications, new vices, like radicalization, in relation to internet usage have also been introduced that were not considered in the internet usage and accessibility campaign. Online radicalization which is a process by which internet users are brainwashed to accept extreme religious, social or political ideologies. Radicalization therefore led to other internet vices like cybercrime, child trafficking and even hacking among others.

Due to internet development and the increase in the number of users, recommender systems for web applications were introduced to anticipate users' preferences in terms of content and information based on their personalized content. This is based on the users' interaction with the system such that the information is analyzed over time to determine the preferences. Preferences could also be determined based on similarities discovered between preferences from different user groups (community opinion) where the trend could also be used as the basis for recommendation. Recommender systems are mainly based on two categories; collaborative and content-based filtering methods.

Using this idea, internet menace could also be curbed by being able to understand users' online activities. Radicalization does not happen overnight, and therefore with accumulated information about user's browsing activities, the system would be able to identify and relate the trend that leads to radicalization. The relationships between websites and users behavior is analyzed in order to identify the trend in regard to internet usage. The importance of websites to users will be considered based on time spent, frequency of accessibility to determine the probability of radicalization using Bayesian theory and association rule mining methods. The following sections will review what has been done by other scholars in this field, concentrating on the recommender systems techniques and how they could be applied in curbing the menace associated with the use of web applications.

2. THEORETICAL REVIEW

Users' activities on the website are guided by the guiding principles of ethics. Users tend to misuse online sites knowingly thereby ending up radicalized as a result of a continuous use of some specific online sites. Ethics has been defined in various ways by different scholars. According to Mohd & Sembok [7], these are moral values that help guide human behavior, actions, and options. They can be defined in two ways normative and prescriptive. For the normative description, ethics are considered to be well-based standards of right and wrong that suggest what humans should to do,

usually in terms of rights, obligations, benefits to society, fairness, and specific virtues. Ethics also involves values that related to virtues of honesty, compassion, and loyalty. For the prescriptive description, ethics refers to the study and nurturing of personal ethical standards, as well as community ethics, in terms of behaviour, feelings, laws, and social habits and norms which can deviate from more universal ethical standards [7].

Hilty [5] also defines the term ethics in two ways; philosophical reflection and the other to more practical governing principles, which is the principle of moral conduct governing an individual or a group.

Conclusion from the above definitions can therefore be derived that, ethics are standards of right or wrong as perceived by human being in the use of technology. What is perceived as wrong in the human eye could remain wrong as long as the person in question perceive it so. It does not have to be controlled by government institutions laws or standards for them to be seen so.

In relation to technology, and more so internet, there are many issues that come as a result of development in technology. These could be good or bad depending on the situation and information use. Social sites like Myspace, Facebook, Friendster, Twitter always enable users to learn something new. In Friendster, for example, students can share comments and pictures as well as ideas which could be beneficial or negative depending on the perception of the user. Negative influence such as addiction to pornographic, radicalization and even exposure to fraud [12]. Ethical issues in Information Technology (IT) involves three "rights" which are;

- Right to know: This is the extent to which information can be accessed from IT infrastructure.
- Property right: This is the right to that IT facilities are protected. For example, users can protect their computers from viruses
- Right to privacy: This refers to right to user's privacy. Every technology user should be responsible to protect their information like passwords from being used by other people.

As much as these rights guarantee freedom of information access and protection, some people, using skills acquired in a bid to enforce these rights, violate them. They therefore ends up in fraud, radicalization or even hacking.

Other challenges that face the industry, especially in academic institutions is the homogeneous society that makes it impossible to come up with umbrella technology ethical standards that can be used across board. In higher learning institutions, there are many people with different age groups, class, gender and affiliations. People of different occupations, age groups and educational backgrounds who live in different countries representing a variety of cultures use resources available through the global computer network. This, however, complicates the problem of developing universal standards of behaviour and a system of ethical norms, which could be widely recognized in the World Wide Web. Meanwhile many believe that such a system is really needed today. This is as mentioned by Gordon & Johnson [4]. However, this heterogeneous society brings along vices that are shared among the users thereby ending up affecting them in the long run.

According to Hilty [5], much has been done in terms of enforcing IT ethics. The adoption of the "ACM Code of Ethics and Professional Conduct" by the ACM Council in 1992. Also, IFIP was introduced SIG 9.2.2, its Special Interest Group on a Framework for Ethics of Computing. He also mentioned that, there have been many challenges for IT ethics for over a decade. For example in privacy protection, data security, trust in complex systems, online communication, freedom of speech, intellectual property and sustainable development

As much as technology was introduced in learning institutions over a decade ago, concentration has been on the application of technology and less on ethics and influence effects. As mentioned by Mohd & Sembok [7], technology ethics have been negatively been affected mainly because of lack of awareness of technology information security issues, the rapidly evolving complexity of systems that are difficult to implement and operate, the ease at which one can access and reach information and communication technology, the anonymity provided by these technologies, and the transnational nature of communication networks. Cybercrime and online radicalization have been issues of the global ethical aspects that have not been resolved fully despite government regulations and policies. Even where awareness is growing and where legislation may be adequate, capacity to use information security technologies and related procedures, as well as to protect against, detect and respond effectively, to cybercrime and online radicalization, is low.

Globally, internet and communications cut across territorial boundaries, thereby creating a new realm of human activities, and undermining the feasibility and legitimacy of applying laws based on geographic boundaries. The new boundaries, which are manifested in the monitor screens, firewalls, passwords, intruder detection, and virus busters, have created new personalities, groups, organizations, and other new forms of social, economic, and political groupings in the cyber world of bits. Traditional border-based law making and law enforcing authorities find this new environment of cyber boundaries very challenging.

The following are some of the main malpractice issues that have not been adequately addressed by the current measures in place, as mentioned by Mohd & Sembok [7];

Cybercrime: Research conducted by Computer Security Institute (CSI, 2003) in Asia-Pacific region indicate that the threat from computer crime and information security breaches continues persistently thereby resulting to financial losses 50 – 60 % between 2002 to 2004.

Pornography: Internet brings great benefits, but also enormous risks to children. Among the worst forms of cybercrime is the global trade in child pornography. International criminal investigation have been trying to handle this situation but with a very small margin.

Piracy: In early 2002, International Planning and Research Corporation (IPR) completed an analysis of software piracy for the year 2001 as part of an ongoing study for the Business Software Alliance (BSA) and its member companies. The world piracy rate increased in two consecutive years, 2000 and 2001. The 2001 piracy rate of 40% is a marked increase from 37% in 2000. Both years were up from the low set in

1999 at 36%. Since then, the rate of piracy has been increasing tremendously over the years.

Neumann & Stevens [10] also mentioned online radicalization as the one of the biggest challenge that security institutions have not been able to handle. They define radicalization as "the process (or processes) whereby individuals or groups come to approve of and (ultimately) participate in the use of violence for political aims". Most policymakers and scholars have only a general idea that internet is the main breeding ground for radicalization but only the most cursory idea of how it works.

Therefore, online platform forms an ideal cheap ground for extremist organizations to sell their ideals. According to Neumann & Stevens [10], internet can be used in three main ways by the extremists;

- To illustrate, communicate their ideological messages and/or narratives
- Internet provides a risk-free platform for potential recruits to find like-minded people. It is also used to for a network of these like-minded
- It creates a new social platform in which unacceptable views and behavior are normalized which would have otherwise been rejected. "Surrounded by other radicals, the internet becomes a virtual 'echo chamber' in which the most extreme ideas and suggestions receive the most encouragement and support."

According to Olumoye & Y [11], he suggested the following measures that could curb negative effects of technology:

- The society must ensure each person is accountable for everything he or she does, no matter how inexplicable his or her action may appear.
- Since there are growing complexities of ethical and social issues that revolve around multiple breaches, it becomes imperative for the educators and computer
  professional bodies to develop curriculum on ethical and professionalcodes of conduct in the information society.
- There is need to lay emphases on information systems security controls
- The government should develop a comprehensive laws and legislations to create a sense or awareness of compliance requirements that affects information systems professionals.
- Our law enforcement agents should be more sophisticated in their computer crime investigation. This can been enhanced with the use of computer forensics, which is a formal investigative technique used in evaluating digital information for judicial review

The use of intelligence techniques have of late taken a centre stage especially in the use of web based applications in identifying and categorizing ideas based on users preferences. According to Mukhopadhyay, Vij, & Wanaskar [8] recommender systems are tools used to filter, sort and order items and data based on opinions of users or community to assist users to determine content of interest from a doubtless overwhelming set of decisions. Two algorithms became very popular: collaborative filtering and content-based filtering

Content-based recommender systems work with individual profiling of users from the very beginning. A profile has biographic data about a user and preferences based on items rated by the user. In the entire process of recommendation, the system makes a comparison of items that were already positively rated by user with items that were not rated and looks for similarities. This is carried out mainly using tags or keywords. In this case the profiles of other users are not essential.

Collaborative filtering method involves searching and locating users in a community that share item preferences based on their common online browsing habits. If two or more users have similar or closely similar rated items, they are considered to have similar preferences. A user gets recommendations to choose items that he/she has not rated before, but that were already positively rated by users in his/her neighbourhood [8].

Mukhopadhyay, Vij, & Wanaskar [8] identified the following challenges affecting recommender systems;

1. Cold-start: This occurs when providing recommendations to new users. The system is not able to recommend preferences to new users as their profiles are almost empty. Cold-start also affects new items that have not been captured by the system and haven't been rated before.

2. Trust: Consideration of people with short history cannot be compared with those with rich history in their profiles. The issue of trust arises towards evaluations of a certain customer.

3. Scalability: The number of users continues to grow tremendously and therefore, the system needs more resources for processing information and forming recommendations. Majority of resources is utilized during determination of users' preferences.

4. Sparsity: This refers to the problem of lack of information. In online shops that have a huge amount of users and items there are some users who rate just a few items.

5. Privacy: In order to receive the most accurate and correct recommendation, the system must acquire the most amount of information possible about the user, including demographic data, and data about the location of a particular user. Naturally, the question of reliability, security and confidentiality of the given information arises. Many online shops offer effective protection of privacy of the users by utilizing specialized algorithms and programs.

This idea could be extended in profiling users interest and thereby be in apposition to determine and predict the kind of a person the user is based on his/her online activities. These systems are developed using intelligence techniques such as association rule mining, Bayesian theory, cluster analysis, reinforcement learning among other techniques.

In association rule for example as stated by Mukhopadhyay, Vij, & Wanaskar [8] an upgraded association rule mining method cab be used in recommender systems because of its scalability and gives high precision in results determination. Using this method, the weight of the web page is given in binary form to pages that are visited to find whether the page is present or not. This method assumes that if the web page is visited by the user, that page is considered important specifically to that user. However, not all the pages visited by the user are of interest. Some users may visit a page looking for useful information but it may not have what the user is looking for. Therefore, factors like time spent by the user and visiting frequency of the page are considered during web page

calculation. This idea could also be used in identifying web pages of interest to the user in order to understand users browsing behaviour thereby predicting their personalities. This therefore could be used to curb vices like cybercrime, radicalization among others.

Another technique highlighted by different scholars in recommender systems and data mining, especially in probability theory and statistics is Bayes' theorem also referred to as Bayes' rule. It is a result that is of importance in the mathematical manipulation of conditional probabilities. Bayes rule can be derived from more basic axioms of probability, specifically conditional probability. This has also been extended to machine learning.

According to Tipping [13] when applied, the probabilities involved in Bayes' theorem may have any of a number of probability interpretations. In one of these interpretations, the theorem is used directly as part of a particular approach to statistical inference. ln particular, with the Bayesian interpretation of probability, the theorem expresses how a subjective degree of belief should rationally change to account for evidence: this is Bayesian inference, which is fundamental to Bayesian statistics. However, Bayes' theorem has applications in a wide range of calculations involving probabilities, not just in Bayesian inference.

Mathematically, Bayes' theorem gives the relationship between the probabilities of A and B,P(A) and P(B), and the conditional
probabilities of A given B and B given A, P(A|B) and P(B|A). In its most common form, it is:

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}.$$

The meaning of this statement depends on the interpretation of probability ascribed to the Bayesian interpretation [13].

Neal [9] provides the following steps in implementation of Bayesian method;

1. Formulation of knowledge about the situation probabilistically
2. Gathering of data
3. Computing the posterior probability distribution for the parameters
4. Given the observed data
5. Use of the identified posterior distribution to:
   i. Reach scientific conclusions, properly accounting for uncertainty.
   ii. Make predictions by averaging over the posterior distribution.
   iii. Make decisions so as to minimize posterior expected loss.

The posterior distribution for the model parameters given the observed data is found by combining the prior distribution with the likelihood for the parameters given the data. This is done using Bayes' Rule:

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}.$$

Another form of Bayes's Theorem that is generally encountered when looking at two competing statements or hypotheses is:

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B|A)P(A) + P(B|\neg A)P(\neg A)}.$$

For proposition A and evidence or background B,

- P(A),the prior probability, is the initial degree of belief in A.
- P(-A), is the corresponding probability of the initial degree of belief against A: 1-P(A)=P(-A)
  - P(B|A), the conditional probability or likelihood, is the degree of belief in B, given that the proposition A is true.
  - P(B|-A), the conditional probability or likelihood, is the degree of belief in B, given that the proposition A is false.
  - P(A|B), the posterior probability, is the probability for A after taking into account B for and against A.

Tipping [13], has pointed out the main advantage and disadvantage of the Bayesian method. The greatest advantage of a Bayesian approach is that, there is an automatic preference for simple models that sufficiently explain the data without unnecessary complexity. This property holds even if the prior P(W) is completely uninformative. The practical disadvantage of the Bayesian approach is that it requires the modeller to perform integrations over variables, and many of these computations are analytically intractable. As a result, much contemporary research in Bayesian approaches to machine learning relies on, or is directly concerned with, approximation techniques.

The following are other artificial intelligence techniques as identified by Wu et al. [14]

a) Decision tree learning

This method uses a decision tree as a predictive technique mapping observations about an item to conclusions about the item's target value. This techniques is among the most predictive modelling technique used in statistics, data mining and machine learning. This approach depicts a tree like structure with leaves representing class labels and branches conjunctions of features that lead to class labels.

In decision analysis, a decision tree can be used to visually and explicitly represent decisions and decision making. Decision tree learning is a method commonly used in data mining. The goal is to create a model that predicts the value of a target variable based on several input variables.

Data comes in records of the form:

$$(\mathbf{x}, Y) = (x_1, x_2, x_3, ..., x_k, Y)$$

The dependent variable, Y, is the target variable that we are trying to understand, classify or generalize. The vector x is composed of the input variables, x_1, x_2, x_3 etc., that are used for that task.

This method however has got the following limitations;

- Practical decision-tree learning algorithms are based on heuristics such as the greedy algorithm where locally-optimal decisions are made at each node. Such

algorithms cannot guarantee to return the globally-optimal decision tree.

- Decision-tree learners can create over-complex trees that do not generalize well from the training data (over fitting).
- There are concepts that are hard to learn because decision trees do not express them easily, such as parity (the evenness or oddness of the number of bits with value one within a given set of bits, and is thus determined by the value of all the bits. It can be calculated via an XOR sum of the bits, yielding 0 for even parity and 1 for odd parity. This property of being dependent upon all the bits and changing value if any one bit changes allow for its use in error detection schemes.) or multiplexer (a device that selects one of several analog or digital input signals and forwards the selected input into a single line) problems. In such cases, the decision tree becomes prohibitively large.
- For data including categorical variables with different numbers of levels, information gain in decision trees is biased in favour of those attributes with more levels.

b) Association rule mining

This is a popular and well researched method for discovering interesting relations between variables in large databases. It is intended to identify strong rules discovered in databases using different measures of interestingness. Based on the concept of strong rules, Agrawal & Srikant [2], introduced association rules for discovering regularities between products in large-scale transaction data recorded by point-of-sale (POS) systems in supermarkets. For example, the rule $\{onions, potatoes\} \Rightarrow \{burger\}$ found in the sales data of a supermarket would indicate that if a customer buys onions and potatoes together, he or she is likely to also buy hamburger meat. Such information can be used as the basis for decisions about marketing activities such as, e.g., promotional pricing or product placements.

As stated by Mukhopadhyay, Vij, & Wanaskar [8], association rule mining technique can be easily used in recommendation systems and it is scalable, gives high precision , and only gives binary weight to the pages that are visited i.e. to find whether the page is present or not. User may visit a page but it may not have useful information for him. So factors like time spent by the user and visiting frequency of the page should be considered for the page consideration. So in association rule mining method the weight of the page is also included.

One limitation of the standard approach to discovering associations is that by searching massive numbers of possible associations to look for collections of items that appear to be associated, there is a large risk of finding many false associations. These are collections of items that co-occur with unexpected frequency in the data, but only do so by chance.

c) Cluster analysis

Also known as clustering is the task of grouping a set of objects so that the objects in the same cluster are more similar to each other than to those in other clusters. It is a main task of exploratory data mining, and a common technique for statistical data analysis, used in many fields, including machine learning.

Popular notions of clusters include groups with small distances among the cluster members, dense areas of the data space, intervals or particular statistical distributions. Cluster analysis as such is not an automatic task, but an iterative process of knowledge discovery or interactive multi-objective optimization that involves trial and failure. It will often be necessary to modify data pre-processing and model parameters until the result achieves the desired properties.

d) Reinforcement learning

As emphasized by Peters et al. [6] Reinforcement learning is an area of machine learning inspired by behaviorist, concerned with how software agents ought to take actions in an environment so as to maximize some notion of cumulative reward. Reinforcement learning is mostly applied in dynamic programming and machine learning.

Reinforcement learning is suitable to problems which include a long-term versus short-term reward trade-off. Reinforcement learning has been applied successfully to problems like, robot control, elevator, scheduling, telecommunications, backgammon and checkers.

Reinforcement learning is considered powerful because of two components: The use of samples to optimize performance and the use of function approximation to deal with large environments. Peters et al. [6] also identify situations where reinforcement learning can also be used successfully. These situations are:

- A model of the environment is known, but an analytic solution is not available;
- Only a simulation model of the environment is given (the subject of simulation-based optimization);
- The only way to collect information about the environment is by interacting with it.

The main limitation in this technique is that, although finite-time performance bounds appeared for many algorithms in the recent years, these bounds are expected to be rather loose and thus more work is needed to better understand the relative advantages, as well as the limitations of these algorithms.

The researcher intends to use the Bayesian theory combined with association rule learning to generate the proposed algorithm. The aspects of recommender systems, using content-based approach will also be considered in predicting what sites are most likely to be visited by the user. This will ensure that the limitations experienced from both methods are covered thereby increasing dependability percentage of the algorithm. In Bayesian technique, much contemporary research to machine learning relies on, or is directly concerned with approximation techniques and is based on

approximation. Since both methods are based on probabilities, combining both increases the chances of acquiring the correct results.

3. DISCUSSION

Machine learning/artificial intelligence is a relatively new area in information technology. A variety of techniques have been introduced each with its benefits and shortcomings based on different applications. More research has been carried out on integration and enhancement of these techniques to form hybrids in a bid to make them much better and effective.

Inspired by Forsati & Rahbar [3] using weighted association mining method; the following criteria are considered;

1. Page duration
2. Frequency of access
3. Indegree – in this case it refers to the average number of objects in a page

These three criteria are used to generate the average weight of the web page. The more the weight, the more important the page is considered by the user.

The associated method is used in capturing the time and webpage access frequency which forms the basis of importance estimation. Generally, a user is assumed to spend more time on more useful web pages than on less useful ones. However, this alone cannot be sufficient as other factors like size of the page and number of objects in a page might equally be important. The formula of duration is given in *Equation (1)* below. Frequency is the number of times that a web page is accessed by a user. The general assumption is that web pages with a higher frequency are of stronger interest to users. However, it is important to consider in the calculating the frequency of a web page the in-degree of that page, in this research the number of objects in a page e.g. images, videos, flashy objects etc. The frequency is calculated using the formula in *Equation (2)* below [8].

Bayesian aspect can then be used to capture and predict user's behaviour based on the usage of the social media like Twitter and Facebook. This therefore concentrates on the probabilistic assumptions specifically on social media. These could be measured through comments posted, likes or tags identifying the trends and already known facts in regard to social media usage and radicalization statistics. In particular, this is used to generate the probability Maximum a Posteriori (MAP) as to whether the user is radicalized or not based on his usage statistics of the social media sites. This is also forms part of the equation generated for association rule mining as *Equation (3)*. Therefore, time spent by a user on a web page, frequency of visiting and probability estimation based on social media usage and information, are then used as three crucial aspects in measuring the user's interest on the web page and relationships among different web pages using *Equation (3)* as indicated below.

$$\text{Duration} = \frac{\text{Time spent}}{\text{Size (bits)}} \qquad \textit{Equation (1)}$$

$$\text{Frequency} = \text{Number of visits} * \text{Indegree} \qquad \textit{Equation (2)}$$

$$\text{Probability} = \text{MAP} \qquad \textit{Equation (3)}$$

$$\text{Weight} = \text{Duration} * \text{Frequency} * \text{Probability} \qquad \textit{Equation (4)}$$

4. CONCLUSION

Internet can no longer be ignored and therefore, monitoring measures efficient in controlling users' activities online need to be put in place that are. As much as more research has been carried out in enhancing artificial intelligence techniques that are used in profiling and recommender systems, more is still required. In this survey, association rule mining has been enhanced to weighted association rule mining considering the average weight of the web page. However, not all web pages should be treated the same way. Social media web pages have different kind of information and should be treated as such, differently from other web pages. Therefore, as much as weighted association rule mining method would be appropriate for other informational web pages, Bayesian theory would be appropriate for social media sites. The combination of these two algorithms would ultimately form an enhanced algorithm with accurate and reliable results.

5. REFERENCES

[1]. Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. San Jose, CA 95120: IBM Almaden Research Center 650 Harry Road.

[2]. Forsati, R. M., & Rahbar, A. (2009). An efficient algorithm for web recommendation systems. AICCSA 2009. IEEE/ACS International Conference on Computer Systems and Applications, 579-586.

[3]. Gordon, L., & Johnson, D. G. (2004). Ethical psychological and societal problems of the application of technologys in education. Unesco Institute For Information Technologies in Education.

[4]. Hilty, L. M. (2002, December). The role of ethics in the field of information and communication technologies. Discussion paper prepared for the technology working group of the Swiss Academy of technical (SATW).

[5]. Peters, J., S, V., & S, S. (2003). Reinforcement learning for humanoid robotics. Third IEEE-RAS interantional conference on Humanoid Robots, 29-30.

[6]. Mohd, T., & Sembok, T. (2003). Ethics of information communication technology (technology). University Kebangsaan Malasia for the regional meeting on ethics of science and technology.

[7]. Mukhopadhyay, D., Vij, S. R., & Wanaskar, U. H. (2013). A hybrid web recommendation system

based on the improved association rule mining algorithm.

[8]. Neal, R. M. (2004). Bayesian methods for machine learning. University of Toronto.

[9]. Neumann, R. P., & Stevens, T. (2009). Countering online radicalization. A strategy for action; the international centre for the study of radicalization and political violence (ICSR).

[10].Olumoye, & Y, M. (2013, November 11). Ethics and social impact of information systems in our society: analysis and recommendations. International journal of science and research (IJSR), 2, 2319-7064.

[11].Rikowski, R. (2009). Using information technology, teaching ethical issues in IT (7th ed.). London south Bank University: Mc. Graw Hill.

[12].Tipping, M. E. (2006). Bayesian Inference: an introduction to principles and practice in machine learning. Cambridge, UK: Microsoft Research.

[13].Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., . . . Zhou, Z.-H. (2007). Top 10 algorithms in data mining. London: Springer-Verlag Limited.

Educational Game "Saringman" to Familiarize Hygiene to Children Since Early Stage

Fiftin Noviyanto
Informatics Department
Universitas Ahmad Dahlan
Yogyakarta, Indonesia

Yana Hendriana
Informatics Department
Universitas Ahmad Dahlan
Yogyakarta, Indonesia

Bayu Nadya Kusuma
Informatics Department
Universitas Ahmad Dahlan
Yogyakarta, Indonesia

Abstract: Learning materials on the environment for children aged growth is quite difficult taught, with the lack of education, teaching materials and appropriate materials where children will easily absorb the material that accompanied the game. Delivery of content has not been used as a medium of learning tools. With the learning media in the form of educational games are expected to help the children understand the importance of keeping the environment clean, especially in terms of waste disposal. The strength of a research conducted, it produced an educational game app that contains materials about the importance of protecting the environment, especially in terms of taking out the trash for children between the ages of growth, which has been tested using black box test and alpha test. Based on the test results and Alpha Black Box test produces gaming applications that can helping the children to understand the importance of protecting the environment, with a view of the game that is user friendly, attractive, lightweight, has a weight of education and can be easily deployed to support learning about the importance of maintaining cleanliness environment.

Keywords: educational games; hygiene; environment

1. INTRODUCTION

More people realize that education is something that is very important for the progress of a nation and the State. With the proper education can impact on the development of technology, along with the growing impact of technology in the learning process, together with the impact on the learning process of education patterns also impact on human prosperity.

Environmental education is one of minimal education science applied to the child's age and growth. Therefore, to create a healthy environment in need of development in terms of non-formal education on solid and sustainable environment. To achieve such a case there should be an understanding of developments in technology and the sciences are ongoing in areas of existing application, the necessary role of academics to contribute to produce innovations of modern learning in addition to the conventional model of learning is done through a game.

Games are learning directly to the pattern of learning by doing. In a game there are a consequence of the player's game to get through the challenges that exist in a game play is run. Lessons learned patterns derived from failures that have been experienced by the player, thus encouraging not to repeat the failure next stage. From the pattern that developed the game, the player will be required to make the learning process independently, digging for information to enrich knowledge of strategy while playing. Thus the game offers a huge potential in building motivation in the learning process.

While only preliminary, results suggest that students experience higher intellectual intensity, intrinsic motivation, and overall student engagement when working in NIU-Torcs compared with traditional approaches to homework and class work in mechanical engineering. By reporting greater levels of challenge and concentration on the one hand, and enjoyment and interest on the other, the experience of students working in NIUTorcs is consistent with concept of engagement and learning through `serious games.' The finding

is also consistent with previous research demonstrating that adolescents report the highest levels of engagement during active leisure activities, especially during games and sports [1].

Circulation of online gaming is increasingly penetrated into the elementary school. This has an impact on their behavior when they are poor school environment. Such concerns are not without reason to remember from year to year the number is not violence by children tends to increase [2].

Regulation of the Minister of Public Works number: 21 / PRT / M / 2006 highlights to achieve a condition of the people who live healthy and prosperous future, it will be necessary the existence of a healthy living environment, aspects of waste, then the word will mean healthy as conditions that would can be achieved if the waste can be managed well so clean from neighborhoods where human activities in it [3].

Environment Workshop (BKKL) which established the trash bank Bantul introducing various appropriate technologies in the field of environmental health, instill love in environmental health sciences for unversity students, encourage community participation in increasing the quality of the field of environmental health [4].

Thus it would need to be built gaming applications "building game education saringman since early for hygiene embed to children" so that children can play at once got a lesson about the importance of protecting the environment, especially in terms of disposing of waste in place.

2. LITERATURE REVIEW

Past research has ever made, entitled "Introduction to Game Development Education Animal Name In English For Elementary Students Learning Media Based Macromedia Flash" will be undertaken by Ghea Fatma Dewi [5]. Object of research in this study is the media making public elemtary school Somokaton, researchers revealed truly students' difficulties in memorizing the sequence of letters in naming the animals, the results of these studies get very decent

validation by experts of material, and with the media in the class or outside classes, students do not feel bored, help in memorizing and can help achieve the goal of learning [6].

The development of information technology, especially the gaming industry which is one of the growing gaming industry is important for brain development, to improve concentration and train properly and solve problems quickly because in the game there are various conflicts which requires us to solve it quickly and precisely [7].

Early Childhood Education (ECE) is very important to do well in a home environment as well as in the educational environment of pre-school. Education pre-school children at age 2 to 3 years is the formation of character, where education is stressed here in the picture, sound and movement combined with an attractive shape and color. At this age children begin to recognize the objects around it. The introduction of culture is also one of the curriculum in pre-school education [8].

The main objectives for developing the website are to educate the community on the benefits of performing the working memory test of the activity of the brain and improvements in social skills and improving poor academic and professional performance, especially in maths and reading comprehension [9]. A large number of literatures have been dedicated to the development of goal modeling, result modeling, ratings and rankings for games prediction [10].

The learning effectiveness and motivational appeal of a computer game for learning computer memory concepts, which was designed according to the curricular objectives and the subject matter of the Greek high school Computer Science (CS) curriculum, as compared to a similar application, encompassing identical learning objectives and content but lacking the gaming aspect [11].

The flow theory is used as a framework to facilitate positive user experience in order to maximize the impact of educational games. Especially, the factors that contribute to flow experience are discussed. The experiential gaming model can be used to design and analyse educational computer games. However, the model works only as a link between educational theory and game design and does not provide the means to a whole game design project [12].

The use of educational games in learning environments is an increasingly relevant trend. The motivational and immersive traits of game-based learning have been deeply studied in the literature, but the systematic design and implementation of educational games remain an elusive topic [13].

2.1. Definition of Rubbish
According to WHO, the rest of the waste is not refrigerate after something end process.

2.2. Definition of Game
Game meaningful "playing", game theory is a way of learning that is used to analyze a number of puz and individuals who demonstrate a rational strategy. The game consists of a set of rules that establish the situation play out of two or more of those groups to maximize their own victory or to minimize the opponent victory.

2.3. Game History
The world begins with the console-game console predecessors such as Atari, Nintendo, Super Nintendo (SNES), and SEGA that displays 2D game that is simple enough. Types of game play include:

1) Arcade / Side Scrolling
2) Racing
3) Fighting
4) Shooting
5) RTS (Real Time Strategy)
6) Role Playing Game
7) Simulation Games
8) Physics puzzle and puzzle

2.4. Flash Game History
Since the flash appears as a media tool for web animation, flash has undergone many evolutions in its development. Originally used for simple animation and minimal interaction, flash began to grow with the addition of Action Script 1.0 which takes the form similar to javascript [14].

In the flash 7 (Flash MX 2004), ActionScript 2.0 was introduced to the public with the addition of variable data types and syntax class.

On flash CS6 (Adobe Flash CS6), ActionScript 3.0 introduced the system embracing OOP (Object Oriented Progamming).

2.5. Understanding Educational Games
Educational game is a game that is designed or disposed of to horsefly stimulate including improved concentration and problem solving. Educational game is one type of media used to provide instruction adds its knowledge through a unique and exciting media.

3. RESEARCH METHODS
3.1 Subject Research
Subjects of this research is to establish the design and implementation of educational games with the game content is problem whose mother about lifestyle issues of protecting the environment by disposing of waste in its place is packed in educational flash game with the hope to increase awareness of children in protecting the environment by not disposing of waste in its place, implemented in programming language ActionScript 3.0 in Adobe Flash CS6.

This research is limited awareness of the importance of instilling an understanding of early childhood in protecting the environment, especially for not littering premises exciting educational game media and educate, then these limits are used in the manufacture of reference educational game. The sample used in this study were children aged 5-11 years.

3.2 Methods Aggregation Data
In the course of this research uses several methods to obtain accurate data, among others:

3.2.1 Methods of Observation
Methods of data collection related to the educational game is done by directly observing a similar game as "Jokowi Save Jakarta".

3.2.2 Methods Library Studies
Book study is the method of data collection is done by looking for information or data relating to research subjects educational games on the books, papers, related to the subject of research.

3.3 System Design Waterfall Model
3.3.1 Needs Analysis Software

Research methods of the game "SaringMan" using this type of research development (or engineered) where the concept of the system is represented by the application of interactive learning systems to improve children's education services to understand the environment.

3.3.2 Systems and Software Design
Saringman educational game application design, images, text, animation, and navigation in the form of software adobe flash circuitry, for character and entity elements drawn using CorelDraw X5, Adobe Illustrator CS6 and Photoshop CS6. Dubbing can be taken from the internet, windows movie maker, and android applications.

3.3.3 Generalization Code
Finding design that has been changed into program codes form.

3.3.4 Testing
Units are integrated and tested game into a complete system to ensure that the software requirements have been met.

3.3.5 Maintenance
The system is installed and used Maintenance includes rectification of errors that are not found in the previous step using the backlog.

3.3.6 Testing Validation
3.3.6.1 Black Box Test
Testing the system reliability in presenting information.

3.3.6.2 Alpha Test
Looking for some respondents were asked members feedback on the format of the level of hospitality applications.

4. RESULTS AND DISCUSSION
4.1 Needs Analysis
This research conducted, obtained the data necessary in the design of educational games using the waterfall method. These data are processed and customized into data that is ready to be implemented in a system.

Applications built has inter specifications:
4.1.1 The material is presented in a variety of text, animation and equipped with games to facilitate the delivery of materials and useful as an attraction decoy.

4.1.2 Game as the culmination of the material submitted by the actions in accordance with the input is entered.

4.2 System Functional Requirements
In making the application program saringman game needed an application that is able to perform functions, among others:

4.2.1 Saringman
- Having a throwing direction that can rotate 180 degrees.
- Having power brightener.
- Power lightening ejected by pressing the left click on the mouse.

4.2.2 Enemy (tongpah)
- The enemy carrying garbage
- Enemies affected by the power illuminator will change for the better.

4.2.3 Strength brightener / Lightening Power

- The power used saringman to sensitize the enemy so as not to pollute the environment.
- The power of lightening in the limit as much as 10 tosses.
- The power of lightening = 0, the game is over.

4.2.4 Level
- There are 3 levels in the game
- Conditions of entry level 2 is defeated enemy at the first level.
- Conditions of entry level 3 managed to defeat the enemy in the second level
- Terms completing level 3, managed to defeat the enemy is hiding under a rock.

4.3 Design System
4.3.1 Design Concept
CAI concept models using simulation models and models of instructional games. Well-designed programs can motivate children in environmental education.

4.3.2 Design Menu
Facilitate the use of educational game app "saringman". It is necessary to design that divides the content into three levels. This application will be preceded by an intro to see the initial appearance before heading to the main menu. The design of the menu, among others:
- Intro
- Menu Utama
- New Game
  - Animasi
  - Main Level 1
  - Main Level 2
  - Main Level 3
- Credit
- Help
- Exit

4.3.3 Graphic Design
In the manufacture of button design, character, and background is divided into 3 phases:
4.3.3.1 Skatch Phase
At this stage of the sketch using software CorelDraw X5 in making the layout of the buttons, character, and background
4.3.3.2 Coloring Phase
On stage staining using Illustrator CS6 software in the coloring of the button, the character, and background
4.3.3.3 Dimensional Effects of Color Phase
At the stage of awarding the dimensions ranging from shadow color and gradation effects using Adobe Photoshop software.

4.4 Implementation Design
4.4.1 Character
4.4.1.1 Saringman

Figure 1 Main Character

Figure 1 shows the main character a good ranger, and care about the environment.

4.4.1.2 Enemy

Figure 2 Enemy Character

Figure 2 shows the antagonist characters, citizens are less concerned who named tongpah.

4.4.1.3 Lightening Power
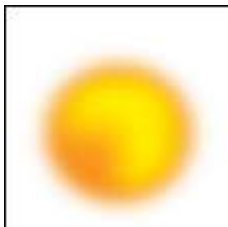
Figure 3 Lighting Power

Figure 3 shows the lightening power that the main character had to resuscitate the enemy character.

4.4.2 System
4.4.2.1 Intro Menu

Figure 4 Intro Menu

Figure 4 shows initial appearance before entering the main menu contains intro introduction of characters and storyline.

4.4.2.2 Main Menu

Figure 5 Main Menu

Figure 5 shows 4 contains the title and menu options that new*game*, credit, Help, and Exit.

4.4.2.3 Level Menu

Figure 6 Level 1

Figure 7 Level 2

Figure 8 Level 3

Figure 6, 7, 8 shows contains the core *game* from level 1 to level 3.

4.4.2.4 Help Menu

Figure 9 Help Menu

Figure 9 shows contains help on how to use the game of saringman.

5. CONCLUSION

Based on the results of the analysis, design and implementation has been done before, it can take several conclusions, among others:

1) This research resulted in a software *(software)* about the game in order to inculcate hygiene education to children from an early age

2) Based on testing with tests conducted *Blackbox method*, this system is appropriate and can be used to help the understanding of consciousness for environmental menjada to children.

3) Testing *alpha test*. From the results of questionnaires that have been carried out concluded that the simulation is quite effective and attractive as well as facilitate the understanding of the importance of protecting the environment.

6. REFERENCES

[1] Coller, B. D., & Shernoff, D. J. (2009). Video game-based education in mechanical engineering: A look at student engagement. International Journal of Engineering Education, 25(2), 308.

[2] Joglo Semar News, June 2013. OnLine Game. pp.7

[3] Faizah, 2008, Household Waste Management Community based (Case Study in Yogyakarta), Tesis S-2 Universitas Dipenogoro Semarang.

[4] Diegenetika, 2013, Garbage Bank Indonesian version, first in Indonesia, http://banksampahbantul.com/news.

[5] Putri, Ghea, 2012, Game Development Education Introduction Animal Name In English For Primary Students Learning Media Based Macromedia Flash, Universitas Nergeri Yogyakarta.

[6] Setiana Nana, 2008, Decision support system based on the work of determining a web-based personality test, Universitas Ahmad Dahlan.

[7] Hendriana, Y., & Ariyana, R. Y. (2015, May). Multimedia adventure game folklore "Doyan Nada" for improving the cultural understanding of Sasak (Lombok) to children. In Intelligent Technology and Its Applications (ISITIA), 2015 International Seminar on (pp. 251-256). IEEE.

[8] Zahrotun, L., Hendriana, Y., & Saputra, D. 2015. Learning Media Introduction of Yogyakarta Culture For Early Childhood 2-3 Years. International Journal of Computer Techniques (IJCT) V2 (5): Page (1-5).

[9] Chiong, R., & Jankovic, L. (2008). Learning game strategy design through iterated Prisoner's Dilemma. International Journal of Computer Applications in Technology, 32(3), 216-223.

[10] Uzoma, A. O., & Nwachukwu, E. O. 2015. A Hybrid Prediction System for American NFL Results. International Journal of Computer Applications Technology and Research, 4(01), 42-47.

[11] Papastergiou, M. (2009). Digital game-based learning in high school computer science education: Impact on educational effectiveness and student motivation. Computers & Education, 52(1), 1-12.

[12] Kiili, K. (2005). Digital game-based learning: Towards an experiential gaming model. The Internet and higher education, 8(1), 13-24.

[13] Moreno-Ger, P., Burgos, D., Martínez-Ortiz, I., Sierra, J. L., & Fernández-Manjón, B. (2008). Educational game design for online education. Computers in Human Behavior, 24(6), 2530-2540.

[14] Griffith, Chris. 2010, Real World Flash Game Development. Canada: Publisher ELSEVIER Canada.

Traffic Light Controller System using Optical Flow Estimation

Neha Parashar

Department of Computer Science and Engineeirng

Mewar University

Chittorgarh, India

Shiv Kumar

Department of Computer Science and Engineeirng

Mewar University

Chittorgarh, India

Abstract: As we seen everyday vehicle traffic increases day by day on road is causing many issues. We face many traffic jams due to the inefficient traffic controlling system which is unable to cope up with the current scenario of traffic in our country. To overcome such drastic scenario and looking at current traffic volume we need to develop a system which works on real time processing and works after determining the traffic density and then calculating the best possibility in which the traffic on particular cross road is dissolved. Also, it helps in saving time as on traffic roads. In present traffic control system when there is no traffic on road but the static signal not allow traffic to move to cross and it changes after at fixed interval so at every cycle this amount of time is wasted for unused traffic density road and if one road is at high traffic it continuously grows till human intervention. The basic theme is to control the traffic using static cameras fixed on right side of the road along top of the traffic pole to check the complete traffic density on other side of the road. This system will calculate number of vehicles on the road by moving detection and tracking system developed based on optical flow estimation and green light counter will be based on the calculated number of vehicles on the road.

Keywords: Optical flow estimation, Moving object detection, tracking, Morphological operation, Blob analysis, Camera.

1. INTRODUCTION

In the current scenario of fixed time traffic lights, many a times situation arises wherein there is heavy incoming traffic only from one side of the intersection and the rest are relatively empty. In this case the people on the heavily occupied side have to wait for unreasonably long time as the green light timer is fixed for each side which fails to take in to account that there is no traffic to pass from the other sides. This prolonged waiting time increases the average waiting time of every person in the traffic. Though care is taken while setting these timers by government officials according to the proportionate amount of traffic present on different sides of the intersection but this can never be so flexible as to adapt to the dynamic traffic throughout the day. Moreover, some areas of high traffic volume may receive scanty traffic at some point of the day and some low traffic volume areas might get congested likewise,

which leads to an additional increase in waiting time because these timers are set according to average volume of traffic corresponding to the different areas of the city respectively but not according to the different hours of the day. This fixed nature of the present traffic light timers turns out to be ineffective not only during the day but also during late night hours. At night, when there is negligible traffic present, the timers lead to an unreasonably long waiting time. Additionally most people, seeing empty roads ahead of them, tend to jump lights at this point, which quite often lead to accidents [1].

Our proposed system takes into account all these issues by dynamically changing traffic light timers. It intelligently recognizes the volume of the traffic at each side of the intersection thereby providing an adequate amount of time for the traffic to pass. The system monitors traffic throughout the day and takes care of the inability of the fixed timers to adjust

as per the traffic. Since our system regulates the flow of traffic at night also, it helps in minimizing the chances of an accident.

2. LITERATURE SURVEY

E. Atkociunas, R. Blake, A. Juozapavicius, M. Kazimianec (2005) presented paper on "Image processing in road traffic analysis". In this Researches and developments have been performed in image a processing technique which is applied to traffic data collection and analysis [2].

Pejman Niksaz (2012) has proposed a paper "Automatic traffic estimation using image Processing" that estimates the size of traffic in highways by using image processing has been proposed and as a result a message is shown to inform the number of cars in highway [3].

P.D. Kamble, S.P. Untawale and S.B. Sahare (2012) presented paper on "Application of image processing for traffic queue length" to measure queue parameters accurately, traffic queue algorithm is used [4].

Raad Ahmed Hadi, Ghazali Sulong and Loay Edwar George (2014) presented paper on "Vehicle Detection and Tracking Techniques: A Concise Review" in which they present a concise overview of image processing methods and analysis tools which used in building these previous mentioned applications that involved developing traffic surveillance systems [5].

 Ramesh Navi, Aruna M. G. (2014) presented paper on "Traffic Event Detection using Computer Vision" they proposed a system in which process includes the subtasks of data collection, negative positive separation, creating training samples, creating description files, haar training and a strong action program which can detect vehicles and the traffic events [6].

Pallavi Choudekar, Sayanti Banerjee, Prof. M.K.Muju presented paper on "Real Time Traffic Control using Image Processing" propose a system for controlling the traffic light by image processing. The system will detect vehicles through images instead of using electronic sensors embedded in the pavement [7].

Overall, the references have advantages such as reporting of speed violation, traffic congestion, accidents, low cost and setup with good accuracy and speed. Some of the disadvantages occurring are variation of ambient light, 3D images are not

supportive, and it is difficult to detect vehicle features in windy and other weather conditions.

3. PROBLEM STATEMENT

Most of Traffic Controllers in India are Manual Controlling and Fixed Time Traffic Signalization. They lead to traffic accidents at intersections. These crashes are caused by drivers' frustration because of long intersection delays. Without police or other forms of enforcement, the long delays have led to road crashes caused by drivers' disobedience of traffic signals and dangerous driving maneuvers in an effort to beat the signal so as to avoid the long delay. This is owing to the fact that some drivers will not wait for the green times when there are no vehicles approaching the intersection, thus Red Light Running (RLR) will occur. In addition, drivers will try to change lane so as to be close to the stop line. The fuel emission and environmental aspects have an influence on people living near the intersections.

4. OBJECTIVE

1. To minimize estimated vehicle delays and stops.
2. To maximize intersection capacity.
3. To design automate control system for traffic on streets.

5. PROPOSED SYSTEM

A scene should be selected from a static camera. Our test movies are selected from urban surveillance videos. Some pre-processing operations have to be done for making the scene ready to process. After that propose algorithm apply on video show below:

Video Frame Acquisition
↓
RGB to Intensity
↓
Mean Calculation
↓
Optical Flow Estimation
↓
Median Filter
↓
Morphological Close

756

```
┌─────────────────────────────────────┐
│           Blob Analysis             │
└─────────────────────────────────────┘
                  │
                  ▼
┌─────────────────────────────────────┐
│       Tracked Detected Objects      │
└─────────────────────────────────────┘
                  │
                  ▼
┌─────────────────────────────────────┐
│        Count Number of Vehicle      │
└─────────────────────────────────────┘
                  │
                  ▼
┌─────────────────────────────────────┐
│         Traffic Light Control       │
│ (Control on the basis of number of  │
│             vehicle)                │
└─────────────────────────────────────┘
```
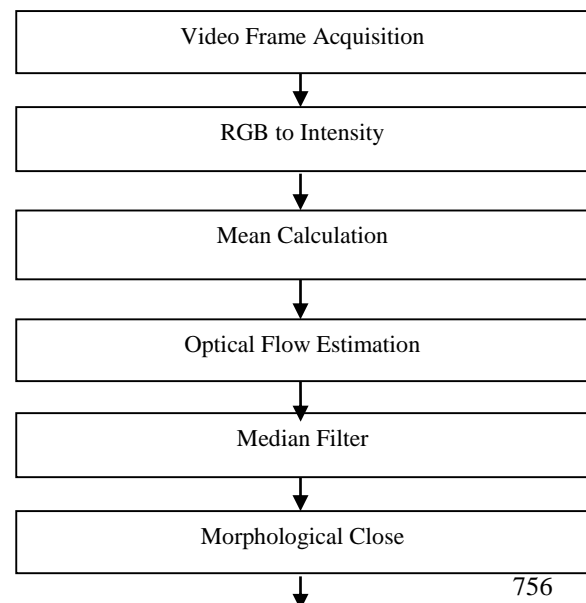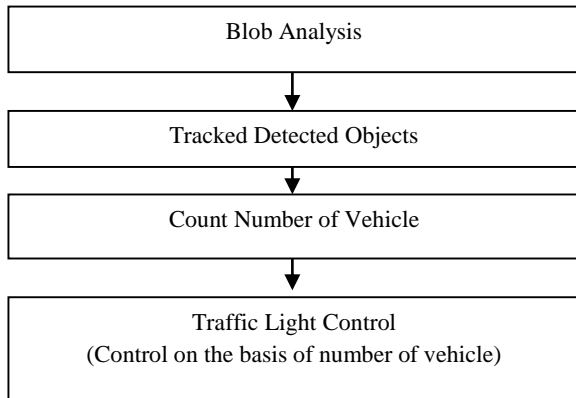
Rules for green light signaling & time control:

Rule 1: If

Number of vehicle < 0

      Then green light counter = 5 sec

Rule 2: If

Number of vehicle < 15

      Then green light counter = 10 sec

Rule 3: If

Number of vehicle < 35

      Then green light counter = 30 sec

Else

      green light counter = 60 sec

In proposed system we set maximum time counter to 60 sec and minimum time counter to 5 sec. Here we show three type of case. We decide the green light counter according to distance and time. In how much time a particular distance travel by vehicle. The counter time depends on distance because the distance at every intersection road is different.

6. THEORETICAL WORK

To design advance traffic control system firstly we setup a camera at 11 feet height. Record video from the camera and perform multiple operations on it discussed below:

6.1 RGB to Intensity

First step for this project is to convert the video we intend to simulate from RGB to intensity. RGB (red, green, blue) are the three colors that can be mixed to become any other colors in the color spectrum. In an RGB video, each pixel is represented by a combination of these colors. Though it provides a more accurate visual representation of the recorded object(s), having to detect 3 colors in every pixel is redundant. Hence, the simplest way is to convert the video from RGB to intensity. What converting the video to intensity does is represent each pixel in the video with a value ranging from 0 to 255. 0 being the color black; 255 being the color white. Any values in-between are shades of gray.

6.2 Mean Calculation

The mean of every frame is calculated on gray-scale format.

6.3 Optical Flow Estimation

Optical flow has been used in conventional video surveillance systems to detect motion, but the purpose of using optical flow in such systems is just to detect moving objects. However, the purpose of using optical flow in our system is to provide statistical traffic flow information. The role of optical flow in our system is essential and critical to final performance. Thus, application-specific assumptions must be considered to choose an algorithm for optical flow estimation from among the many available optical flow algorithms [10].

The optical flow method tries to calculate the motion between two image frames which are taken at times t and t+δt at every position. These methods are called differential since they are based on local Taylor Series approximation of the image signal; that is, they use partial derivatives with respect to the spatial and temporal coordinates.

There are two most used methods, namely:

- Lucas-Kanade
- Horn-Schunck

To solve the optical flow constraint equation for u and v, the Lucas-Kanade method divides the original image into smaller sections and assumes a constant velocity in each section. Then, it performs a weighted least-square fit of the optical flow constraint equation to a constant model for in each section, Ω, by minimizing the following equation:

$$\sum_{x \in \Omega} W^2 [I_x u + I_y v + I_t]^2$$

6.4 Median Filter

In signal processing, it is often desirable to be able to perform some kind of noise reduction on an image or signal. The median filter is a nonlinear digital filtering technique, often used to remove noise. Median filtering is very widely used in digital image processing because, under certain conditions, it preserves edges while removing noise [11].

6.5 Morphological Close

Morphological operations are performed to extract significant features from images that are useful in the representation and description of the shapes in the region; mostly used in image segmentation and pattern recognition. In the proposed system we used both morphological close and erode, respectively, to remove portions of the road and unwanted objects. After morphological closing operation, on condition that vehicle's appearance is not destroyed, objects including many small holes and separated pixels may be connect into one big actual vehicle shape.

6.6 Blob Analysis

In the area of computer vision, blob detection refers to visual modules that are aimed at detecting points and/or regions in the image that differ in properties like brightness or color compared to the surrounding.

6.7 Vehicle Detecting, Tracking and Counting

Blob analysis has the functionality to produce many forms of statistics, which is crucial for detecting and tracking. For now, the bounding box option is checked. A bounding box is an M-by-4 matrix of [x y height width] bounding box coordinates, where M represents the number of blobs and [x y] represents the upper left corner of the bounding box. As the blob moves, the bounding box will follow. Tracking vehicles with boundary boxes are counted.

6.8 Calculation for Green Light Timer of Traffic Light Control System

In this section proposed system provide counter time for green signal light according to the number of the vehicles are present in the video. The counter time for green light decide according to the distance between the intersection and the speed of the vehicle.

In proposed system we manage only green light counter only. System, calculate the number of vehicle present in video. According to the number of vehicle presented in the video set green light timer. The timer for green light varies as number of vehicle varies in detected video. Green light time predefined in the system for different-different condition of the counted vehicle. According to perfect match condition for detected vehicle found system send the green light counter time.

To decide the different-different green light timer depends upon some parameters. These parameters are: distance and speed. Distance parameter, depends upon the distance which is travelled by the vehicle to cross the intersection area. Distance is different-different for different intersection area. Speed, speed of the waiting vehicles is different from each other and everybody drive their vehicle on different speed. To calculate time of green light signal use a basic formula.

Time = Distance/Speed

Time = sec

Distance = meter

Speed = meter/sec

In our proposed system we calculate time on the basis of the parameter which is measured by us to calculate signal light time. In this system we calculate and measure parameter from the intersection road at Bhilwara. At this intersection road/area the 100 meter distance travelled by vehicle within 10 sec. by average speed. So we take the distance as 100 meter. Speed for the system we take 30 km/hr. This is calculated on the basis of average speed of all vehicles. According to Indian govt. speed limit for all vehicle is decided. On the basis of their speed we take the average speed.

According to given formula time decided for proposed system:

$$\text{Time} = \frac{100 \text{ meter}}{30 \text{ km/hr}}$$

Time = 11 sec.

A vehicle which is within the 100 meter (including intersection road/area) road is easily able to cross the road within 10 sec.

7. RESULT

In this section we show the experimental result using data from traffic video. Before applying optical flow estimation on frame the image RGB converted into intensity (gray). Original video and after intensity and median filter shown in fig 1

(a)

(d)
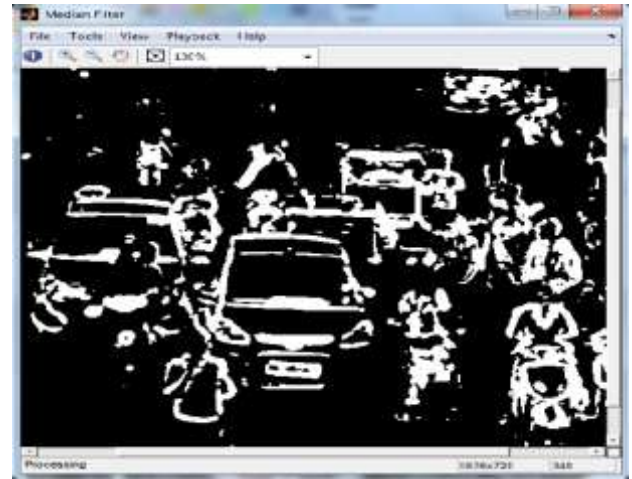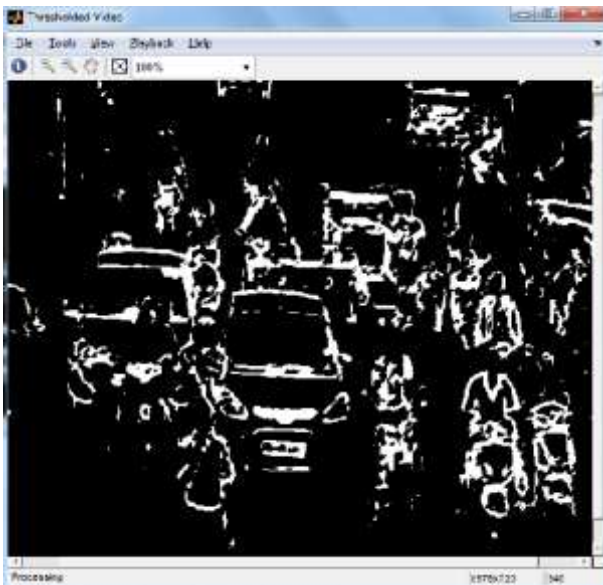
Figure 1: (a) Original Video (b) Intensity (c) Before filter (d) Median filter

The video analyzed for detecting and counting number of vehicle within the video frame. According to number of vehicle it had shown the time period for green signal light. The system was able to produce results for different scenarios.

Case 1: If Number of vehicles is less than 15 then green signal light timer set to 10 sec.

(b)

(c)

Figure 2: No. of vehicle = 8, Green light time = 10 sec

Case 2: If Number of vehicles is less than 30 then green signal light timer set to 30 sec.

Figure 3: No. of vehicle = 18, Green light time = 30 sec

Case 3: If Number of vehicles is more than 30 then green signal light timer set to 60 sec.

Figure 4: No. of vehicle = 35 Green light time = 60 sec

As experimental result shows green light time counter decided according to number of vehicle detected in the video. When the number of vehicles are increase at signal light green light counter also vary.

8. RESULT ANALYSIS

In this chapter we present the test environment and the experimental results analysis of our system.

No of vehicle	Present System timer (sec.)	Proposed System timer (sec.)	Saved Time (sec.)
0	60	5	55
<15	60	10	50
<35	60	30	30
>35	60	60	0

Table 1: Result Analysis

In graphical format result analysis can be shown as given below:
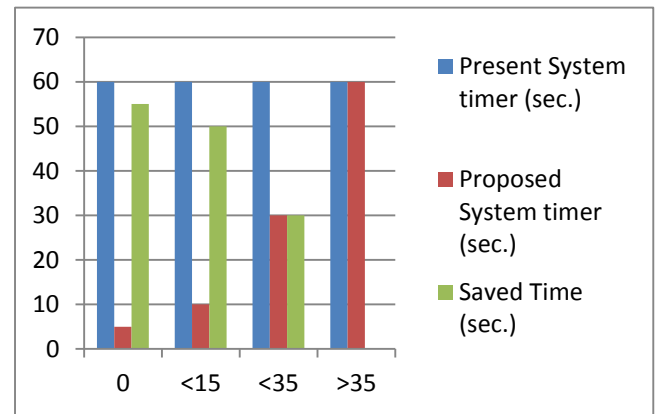
Figure 5: Result Analysis

In present system fuel consumes a lot and it increase air pollution also. According to result analysis it shows we can save a lot of time at traffic signal. Lot of fuel consume by vehicles in waiting time can also be saved. Our proposed system also helps in decrease the fuel consumption by vehicle and decrease the import of fuel from the other countries and saving the foreign currency.

9. CONCLUSION

In this thesis we showed that Control Traffic Light Controller using Optical Flow Estimation. We demonstrate on traffic density estimation and manage traffic signal light based on image processing technique and we successfully calculated the traffic and also manage green light counter. It can be reduce the traffic congestion and time wasted by a green light on an empty road. It shows that it can reduce the traffic congestion and avoids the time being wasted by a green light on an empty road. It is also more consistent in detecting vehicle presence because it uses actual traffic images.

10. FUTURE WORK

Future work will cover complex testing of the system, and more detailed development of modified algorithms. Also the

algorithm and identify overlapping objects transparencies during object tracking.

11. ACKNOWLEDGMENTS

12. REFERENCES

[1.] Pranav Maheshwar, Deepanshu Suneja, Praneet Singh, Yogeshwar Mutneja, "Traffic Image Processing System", Volume 118 – No.2 3, May 2015

[2.] E. Atkociunas, R. Blake, A. Juozapavicius,M. Kazimianec "Image processing in road traffic analysis" Nonlinear Analysis: Modelling and Control, 2005, Vol. 10, No. 4, 315–332.

[3.] Pejman Niksaz, "Automatic traffic estimation using image Processing" International Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 5, No. 4, December, 2012.

[4.] P.D. Kamble, S.P. Untawale and S.B. Sahare, "Application of image processing for traffic queue length" VSRD-MAP, Vol. 2 (5), 2012, 196-205.

[5.] Raad Ahmed Hadi, Ghazali Sulong and Loay Edwar George, "Vehicle Detection and Tracking Techniques: A Concise Review", Vol.5, No.1, February 2014

[6.] Ramesh Navi, Aruna M. G., "Traffic Event Detection using Computer Vision", Volume 16, Issue 3, Ver. II , May-Jun. 2014.

[7.] Pallavi Choudekar, Sayanti Banerjee, "Real Time Traffic Control using Image Processing", Vol. 2 No. 1.

[8.] Joonsoo Lee and Alan C. Bovik, "Estimate and Analysis of urban Traffic flow", 2009.

[9.] Pre Kumar.V, Barath. V, Prashanth.K, "Object Counting and density Calculation using MATLAB".

Detection of Anemia using Fuzzy Logic

Sonu Malu
Mewar University
Chittorgarh, Rajasthan
India

B. L. Pal
Mewar University
Chittorgarh, Rajasthan
India

Shiv Kumar
Mewar University
Chittorgarh, Rajasthan
India

Abstract**:** Medical Science is considered as a field of uncertainty, vagueness and complexity. Fuzzy logic plays an important role to deal with these uncertainty, vagueness and complexity. Detection of diseases in medical is a very difficult task. To improve accuracy rate engineers helping in detection of the diseases by developing the Expert System using Fuzzy Logic. Fuzzy logic consists of many valued logic. It has varying values in the range of 0 and 1 instead of fix values. In this study, we developed a Fuzzy Expert system to detect Anemia on the basis of Symptoms as well as clinical test.

Keywords: **Anemia, Fuzzy Logic, Fuzzy Expert System, CBC Test**

1. INTRODUCTION

The blood that circulates throughout the body performs a number of critical functions. It delivers oxygen, removes carbon dioxide CO_2, and carries life sustaining nutrition's. By acting as the vehicle for long-distance messengers such as hormones, blood helps the various parts of the body communicate with each other. This is carried out by blood cells through working in partnership with the liquid part of the blood (plasma). Anemia is a condition where number of healthy RBC in the blood is lower than normal. It is due to low RBC's, destruction of RBC's or loses of too many RBC's. If your blood does not have enough RBC's, your body doesn't get enough oxygen it needs. As a result you may feel tired and other symptoms. But sometimes it is very difficult to detect anemia on the basis of symptoms only. In the domain of Anemia there is no such boundary between what is healthy and what is diseased. Having so many factors to detect anemia makes doctor's work difficult. So, Experts require an accurate tool that considering these risk factors and give some certain result for uncertain terms.

2. LITERATURE REVIEW

When the studies in the literature related with this classification application are examined, it can be seen that a great variety of methods were used. Among these, [5] Fuzzy System have been used to diagnose the different types of anemia on the basis of symptoms such as Irritability, tachycardia, Memory weakness, Bleeding and Chronic fatigue. Another, [6] diagnose Liver disease using fuzzy logic on the basis of CBC Test which uses 4 parameters such as WBC, HGB, HCT and PLT. [7] Ali.Adeli, Mehdi. Neshat proposed a system to diagnose the heart disease using fuzzy logic. [8] Nidhi Mishra and Dr. P Jha also develop a fuzzy expert system to diagnose the Sickle Cell Anemia.

3. OBJECTIVES

The Objectives are:

1. Detect Anemia using Fuzzy Logic.
2. Classify Anemia on the basis of Accuracy.

4. DESIGN MODEL

4.1 Introduction

Three steps are used to monitor general health and Anemia. But we are focusing only on the Tests and Procedures. Three steps are as follows:

1. Medical and Family Histories
2. Physical Exam
3. Tests and Procedures.

4.2 Design of Fuzzy Logic System

Design model divided into five steps:

1. Problem Specification & define linguistic variables.
2. Define Fuzzy sets.
3. Define Fuzzy Rule.
4. Encode Fuzzy Sets, Fuzzy Rules and Procedures to build Expert System.

5. METHOD

we describe the designing of the fuzzy expert system.

5.1 Design a Fuzzy Logic System

5.1.1 *Problem Specification & Define linguistic variables:* There are 3 input variables and 1 output variables.

Linguistic Variables:

- **For Input Variables**

Table 1 Linguistic Variables for Input Variables

S.No.	Input Variables	Linguistic Variables
1	Hemoglobin	HGB
2	Mean Corpuscular Volume	MCV
3	Mean Corpuscular Hemoglobin Concentration	MCHC

- **For Output Variables**

Table 2 Linguistic Variables for Output Variables

S.No	Output Variables	Linguistic Variables
1	Anemia	Types of Anemia

5.1.2   Define Fuzzy Sets:

- **Input Variables & Value Ranges:**

Table 3 Values for all Input Linguistic Variables[6]

S.No.	Linguistic Variable	Ranges	Values
1	HGB	5 - 13.8 grams/deciliter	Low
		13.9 to 16.3 grams/deciliter	Medium
		16.4 – 18 grams/deciliter	High
2	MCV	60 – 79.9 fl	Low
		79.9 to 100 fl	Medium
		100.1 - 120 fl	High

3	MCHC	28 - 31.9 grams/deciliter	Low
		32 - 36 grams/deciliter	Medium
		36.1 - 40 grams/deciliter	High

- **Output Variables & Value Ranges:**

Table 4 Values for all Output Linguistic Variables[6].

S.No.	Linguistic Variable	Ranges	Values
1.		HGB is 5 – 13.8 g/dl	MicroCytic HypoChromic
		MCV is 60 – 79.9 fl	
		MCHC is 28 – 31.9 g/dl	
2.	Types of Anemia	HGB is 5 – 13.8 g/dl	MicroCytic NormoChromic
		MCV is 60 – 79.9 fl	
		MCHC is 32 - 36g/dl	
3.		HGB is 5 – 13.8 g/dl	MicroCytic HyperChromic
		MCV is 60 – 79.9 fl	
		MCHC is 36.1 - 40 g/dl	
4		HGB is 5 – 13.8 g/dl	NormoCytic HypoChromic
		MCV is 80 - 100 fl	
		MCHC is	

		28 – 31.9 g/dl	
5		HGB is 5 – 13.8 g/dl	NormoCytic NormoCHromic
		MCV is 80 - 100 fl	
		MCHC is 32 - 36 g/dl	
6		HGB is 5 – 13.8 g/dl	NormoCytic HyperChromic
		MCV is 80 - 100 fl	
		MCHC is 36.1 – 40 g/dl	
7		HGB is 5 – 13.8 g/dl	MacroCytic HypoChromic
		MCV is 100.1 – 120 fl	
		MCHC is 28 – 31.9 g/dl	
8		HGB is 5 – 13.8 g/dl	MacroCytic NormoCHromic
		MCV is 100.1 – 120 fl	
		MCHC is 32 - 36g/dl	
9		HGB is 5 – 13.8 g/dl	MacroCytic HyperChromic
		MCV is 100.1 – 120 fl	
		MCHC is 36.1 – 40 g/dl	

5.1.3 Define Fuzzy Rules:

As we have total 3 input variables so total number of possible non conflicting fuzzy inference rules are $3^2 =$ 9 rules.

First 3 rules are for Symptoms based testing:

1. If (irritation is Effective) && (Heart_Rate is High) && (Disorder is cancer) then HGB is low.

2. If (irritation is Effective) && (Heart_Rate is High) &&(Disorder is cancer) && (Blood_Loss is Stomach / intestine bleeding) then HGB is low.

3. If (irritation is Effective) && (Heart_Rate is High) && (Disorder is cancer) && (Blood_Loss is Stomach / intestine bleeding) && (Weak_Memory is Effective) then HGB is low[5].

Further, 3 rules are for the classification of anemia on the basis of MCV only:

4. If (HGB is Low) && (MCV is Low) then MicroCytic is High.

5. If (HGB is Low) && (MCV is Medium) then NormoCytic is high.

6. If (HGB is Low) && (MCV is High) then MacroCytic is high.

At last 9 rules are for the further classification of anemia on the basis of all three parameters such as HGB, MCV, & MCHC.

7. If (HGB is Low) && (MCV is Low) && (MCHC is Low) then MicroCytic is HypoChromic.

8. If (HGB is Low) && (MCV is Low) && (MCHC is Medium) then MicroCytic is NormoChromic.

9. If (HGB is Low) && (MCV is Low) && (MCHC is High) then MicroCytic is HyperChromic.

10. If (HGB is Low) && (MCV is Medium) && (MCHC is Low) then NormoCytic is HypoChromic.

11. If (HGB is Low) && (MCV is Medium) && (MCHC is Medium) then NormoCytic is NormoChromic.

12. If (HGB is Low) && (MCV is Medium) && (MCHC is High) then NormoCytic is HyperChromic.

13. If (HGB is Low) && (MCV is High) && (MCHC is Low) then MacroCytic is HypoChromic.

14. If (HGB is Low) && (MCV is High) && (MCHC is Medium) then MacroCytic is NormoChromic.

15. If (HGB is Low) && (MCV is High) && (MCHC is High) then MacroCytic is HyperChromic[6].

Table 5 Illustration of applied rules with Respect to MF[6]

Rule No.	Linguistic Variable 1	Linguistic Variable2	Linguistic Variable 3	Result

	(HGB)	(MCV)	(MCHC)	
1	Low	Low	Low	MicroCytic is HypoCHromic
2	Low	Low	Medium	MicroCytic is NormoChromic
3	Low	Low	High	MicroCytic is HyperChromic
4	Low	Medium	Low	NormoCytic is HypoChromic
5	Low	Medium	Medium	NormoCytic is NormoChromic
6	Low	Medium	High	NormoCytic is HyperChromic
7	Low	High	Low	MacroCytic is HypoChromic
8	Low	High	Medium	MacroCytic is NormoChromic
9	Low	High	High	MacroCytic is Hyperhromic

5.1.4 Build Fuzzy Expert System:

Form:

Figure 1 Input Form

6. RESULTS AND DISCUSSION

Table 6 Input Values for Results

S. No	Input Variable	Values Ranges	Ranges Selected
1	HGB	10.9 g/dl	5< 10.9 < 18 g/dl
2	MCV	31.00 fl	60< 31.00 < 79.9 fl
3	MCHC	30 g/dl	28< 30 <31.9 g/dl

Figure 2 Result from given Input Values

7. CONCLUSION

In this paper, fuzzy logic is applied to classify and detect Anemia on the basis of CBC Test. The success of fuzzy detection in its application to a real clinical case shows that fuzzy detection is an improvement of probabilistic logic. Results have been shown from this fuzzy expert system with

past time expert system are more efficient and less expensive. It detect anemia on the basis of both Symptoms and CBC Test. From the viewpoint of an end-user, the results of this work can facilitate laboratory work by reducing the time and cost.

8. FUTURE WORK

The future work will focus on developing a machine learning approach to classify different types of anemic RBCs in microscopic images. The method described in this dissertation can be extends in future very efficiently. We can classify anemia on the basis of RBC structure using digital image processing. We can also provide some CBC reports and load that report as it is in our system and detect anemia in future. We can also detect anemia and classify it only on the the basis of CBC Test without using symptoms test.

9. ACKNOWLLDGEMENT

10. REFERENCES

[1] Jameela Ali Alkrimi, Hamid A. Jalab, Loay E. George, Abdul Rahim Ahmad, Azizah Suliman, Karim Al-Jashamy, "Comparative Study Using Weka for Red Blood Cells Classification", World Academy of Science, Engineering and Technology International Journal of Medical, Health, Biomedical and Pharmaceutical Engineering Vol:9, No:1, 2015.

[2] Hematology: The study of Blood.

[3] Yared Alemu, Alemayehu Atomsa, and Zewdneh Sahlemariam, "Hematology", Jimma University, 2006.

[4] National Heart, Lung and Blood Institute, U.S. Department of Health and Human Services, publication no. 11-7629A, September-2011.

[5] Javad Aramideh and Hamed Jelodar, Application of fuzzy logic for presentation of an expert fuzzy system to diagnose Anemia, Indian Journal of science and technology, Vol. 7(7), 933-938, July 2014.

[6] Asma Hashmia, Dr. Muhammad Saleem Khanb, Diagnosis Blood Test for Liver Disease using Fuzzy Logic, International Journal of Sciences : Basic and Applied Research, Volume 20, No 1, pp 151-183, 2015.

[7] N.Tamil Selvi, S. Saranya, P. Usha, M. Yashodha, Perception of Tired Blood and its causes using mining Techniques, International Journal of Engineering Sciences & Research Technology, ISSN: 2277-9655, Vol 4(1), January-2015.

[8] P. Usha, S.Saranya and N. Tamilselvi, Prevalence and risk factors of anemia along with classifiers, International Journal of Scientific & Engineering Research, Vol. 4, Issue 5, May-2013.

[9] Aqueel Ahmed and Shaikh Abdul Hannan, Data Mining Techniques to find out heart diseases : An Overview, International Journal of Innovative Technology and Exploring Engineering, ISSN: 2278-3075, Vol. 1, Issue 4, September-2012.

[10] Eva C. Guinan, Diagnosis and management of Aplastic Anemia, American Society of Hematology, 2011.

[11] G. Licata, Employing fuzzy logic in the diagnosis pf clinical case, Health Journal, Vol. 2, No. 3, 211-224 , 2010.

[12] Mayo Clinic : http://www.mayoclinic.org

Collision Avoidance Protocol for Inter Vehicular Communication

Debasish Roy
Department of Computer Science,
Assam University Silchar-788011, India

Prodipto Das
Department of Computer Science,
Assam University Silchar-788011, India

Abstract: Vehicle to vehicle communication can give us better results by avoiding the major problems in road like collision of vehicles, better route selection in case of traffic congestion, fuel consumption, suitable selection of parking place etc. This paper presents a protocol to avoid the collision of vehicles. High mobility and fast topology changes are the characteristics of Vehicular Ad-hoc Networks (VANETs). To establish the real world environment for VANETs, network simulator NS2 is used. Medium Access Control (MAC) Protocol is used to avoid the collision of transmitted data. The Simulation is done using the proposed Vehicular Ad-hoc On-demand Distance Vector (VAODV) routing protocol, which is a modification of Ad-hoc On-demand Distance Vector (AODV) routing protocol. The proposed VAODV protocol is continuously checks the distance, speed of each vehicle and if it finds that the distance between vehicles is continuously decreasing then in this case it will send a warning textual message to those vehicles that are in accidental situation. Based on this textual information these vehicles will take particular action like vehicle may choose new route if it exists or it may slow down its own speed or it may stop moving by pressing brake. The experimental results are used to find out the performance of VAODV protocol. The performance of VAODV protocol is analyzed with different parameters like end to end delay, throughput, packet delivery ratio, normalized routing load etc.

Keywords: ADC; AES; CAP; CFP; CSMA-CA; MAC; ns-2; OTcl; VAODV; VANETs; Wireless Sensor Networks (WSN)

1. INTRODUCTION

Wireless Sensor Networks (WSN) consist numbers of sensor nodes and they work together to perform a specific task. In WSN base-stations gather the data from all other sensor devices in the networks. Sensor devices are adequate to sense the physical entities. The sensors are equipped with rechargeable batteries [1]. The sensor nodes are communicated through the wireless medium. In WSN usually, all sensor nodes are not reachable to all other sensor nodes. It is a multi-hop network [2]. Components of sensor devices (Fig. 1) are: a sensing and actuation unit (array or single element), a power unit, a processing unit, a communication unit and other application-dependent units [3] [4].
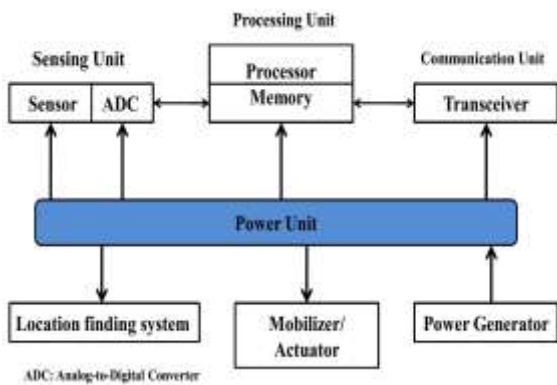
Fig. 1: Components of Sensor Node

Some of the characteristics of wireless sensor networks are topology changes very frequently, sensor devices can join and leave the network at any point of time, sensor devices are having limited power, limited computational capabilities and limited memory, ability to deal with sensor device failure etc.

WSN used in various applications like forest fire detection, flood detection, automated meter reading, vehicle tracking and detection, traffic flow surveillance and many more [4].

1.1 Vehicular Ad-hoc Network

VANETs are special type of mobile ad-hoc networks (MANETs). The basic architecture of vehicular network is shown in Fig. 2.

Fig. 2: Vehicular network architecture

It is a self organizing network which consists of numbers of vehicles. In VANETs, each vehicle is equipped with a device called on-board unit (OBU) which is capable to communicate with other OBUs and road-side units (RSU) using a routing protocol through WSNs. RSU can communicate with the other RSUs. Some RSUs can be used as gateways. Based on these communications VANETs can provide road safety, comfort of driving etc [5] [6] [7].

1.2 IEEE 802.15.4

We have used IEEE 802.15.4 standard MAC protocol in our simulation. IEEE 802.15.4 defines the MAC and physical Layers of wireless sensor networks [8]. It is developed for very low power, low cost, moving devices and for short range wireless communications. It gives support to connect devices are in the field of applications like vehicular, industrial, agricultural, residential and medical sensor etc [9]. The standard is pertinent for self-organizing network along with the essence of network flexibility.

The IEEE 802.15.4 standard MAC sub-layers uses the carrier sense multiple access with collision avoidance (CSMA-CA) mechanism for beacon transmission [10]. There are two modes in IEEE 802.15.4 standard: beaconless and the beacon enabled mode. In beaconless mode it uses the pure CSMA-CA for channel access and works as a basic IEEE 802.11 standard without channel reservations. In beacon enabled mode it uses the hybrid time division multiple access (TDMA). In case of time critical transmission it uses guaranteed time slot(GTS) mechanism to reserve a dedicated time slot within a super-frame [10] [11]. In a Link-layer security is implemented or designed with 128 bit advanced encryption standard (AES) encryption [11]. In beacon enable mode the source periodically sends beacons for sensor nodes synchronization and to establish the personal area networks (PANs). Each super-frame has two parts active and inactive part. Active part is composed of: a beacon, a contention access period (CAP) and a contention free period (CFP) [12]. IEEE 802.15.4 standard is responsible for acknowledging frame delivery, frame validation, beacon management, GTS management, channel access mechanism, supervise device security etc [13].

1.3 AODV Routing protocol

AODV protocol is the advancement of Destination-Sequenced Distance-Vector Routing (DSDV) protocol. AODV is classified as a pure on-demand route acquisition system; nodes do not maintain any routing information or do not exchange any routing table that are not presents in the chosen path[14]. In AODV all the routes are discovered when it is required and are maintained as far as these are being used. Routes are found out through route discovery process. Route discovery starts if a source node required a valid route to some destination node [15].
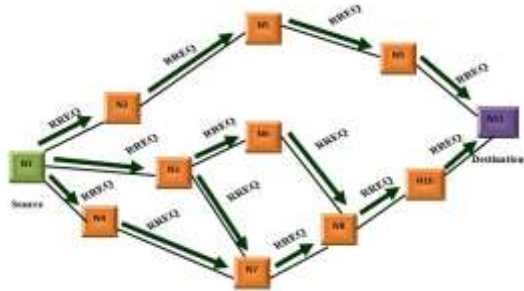
Fig. 3: Route Request Propagation

A route request (RREQ) packet broadcast to all its neighboring nodes by the source node, they again sends the RREQ packet to their neighbors and so on, until destination node found or an intermediate nodes with a route that is fresh enough to reach the destination is found [14] [15] [16]. The propagation of route request is shown in Fig. 3.

AODV utilizes the sequence number that is maintained by all destinations for avoiding routing loops as well as to analyze the newness of routing information [16]. Broadcast ID and node sequence number is maintained by the node itself. For each RREQ made by a node, the Broadcast ID and the IP address are incremented by one for to uniquely identify a RREQ. The source node includes the destination node most recent sequence number and destination IP address, as well as current sequence number of its own and broadcast ID in RREQ [15] [16]. Intermediate node can reply to the RREQ if it has a valid route to reach destination with sequence number greater than or equal to that of the sequence number presents in the RREQ otherwise it rebroadcasts the RREQ [14] [16].
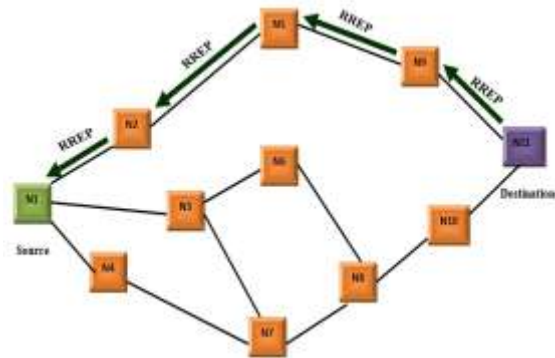
Fig. 4: Route Reply path to the Source

While forwarding the RREQ intermediate nodes record the neighboring node address from where first RREQ packet was received into their own routing table using that information establishes a reverse path to source (Fig. 4). The RREQ packets that are already received if it came later again all these packets are discarded [14] [17]. When the destination gets the RREQ packet or intermediates with fresh enough route, the destination or intermediate replied by unicasting a Route Reply (RREP) control packet back to the neighboring node from which first received the RREQ packet [14] [18]. When the source node received the RREP a complete bidirectional path is establish and the source can start transmission by using this discovered route [14] [16] [17] [18].

1.4 Network Simulator-2 (NS2)

We have used NS2 for simulation of vehicular communication. NS2 is the one of the most popular and efficient research simulator for simulation of WSN. It is an open source, object oriented discrete event-driven simulator which is particularly made for to do research in computer networks. NS2 made of two languages: C++ and Object-oriented Tool Command Language (OTcl), where C++ works as backend scheduler and OTcl interpreter as a frontend scheduler. These two languages are interlinked using the Tcl with classes (TclCL). Simulation internals is defined by the C++ whereas OTcl is used for to schedule discrete events and also to control the simulation by the object gathering and configuring. It supports to design a new protocol, protocol comparison, network traffic analysis etc. A simulation is done using three steps which are: simulation design, configure as

well as to run the simulation and packet tracing which is a text based packet tracing and Network AniMation (NAM) packet tracing [19]. The basic architecture of network simulator is shown in Fig. 5.
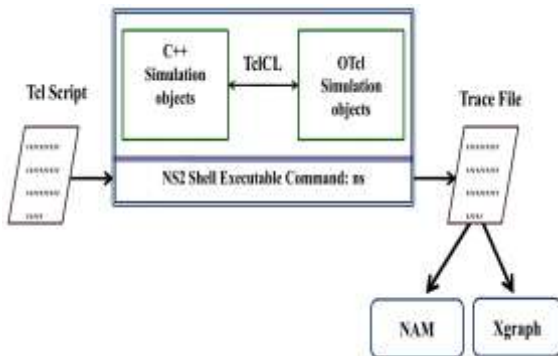
Fig. 5: Basic Network Simulator Architecture

2. PROPOSED ALGORITHM

A collision avoidance algorithm for inter vehicular communication:

Step 1:

Find the co-ordinates of vehicles in milliseconds and store results into the variables X, Y and Z. That means we are storing the location of vehicles. Here we have assumed that the co-ordinate Z is always to be Zero because vehicles are not flying.

Step 2:

If after each millisecond co-ordinates of any one of the vehicle is not matched with the previous co-ordinates value of its own then update the value of the variables X, Y and Z for that vehicle, which means we are storing the new location of that Vehicle.

Step 3:

Calculate the Distance (Eq.1) between two vehicles in each millisecond using the formula:

Let (X1, Y1) be the coordinate of vehicle A and (X2, Y2) be the coordinate of vehicle B, then the distance between vehicle A and vehicle B is

$$Distance = \sqrt{(X2 - X1)^2 + (Y2 - Y1)^2} \qquad Eq.1$$

Step 4:

If the distance between vehicles is less than or equal to warning distance then monitoring vehicle will send a textual warning message to the monitored Vehicle.

Step 5:

If the receiver receives the message, then the receiver will take action according to the received textual warning message and will send an acknowledgement message to the sender.

Step 6:

After certain wait if the sender has not got any response from the receiver then repeat Step 4.

3. IMPLEMENTATION DETAILS

We have considered four different scenarios for simulation and these are:

3.1 Scenario One

Fig. 6: One source one destination moving towards each other

In this case (Fig. 6), both of the vehicles A and B are moving towards each other at a speed of 100 km/h. Initially distance between them is 300 meters. Vehicle B is the source node and vehicle A is the destination node. Vehicle B is monitoring its distance from vehicle A continuously and the moment it found that the distance monitored is less than or equal to the warning distance it will immediately send a warning textual message to the destination vehicle A. The vehicle A will take action according to the received textual message and gives a response to the sender.

3.2 Scenario Two
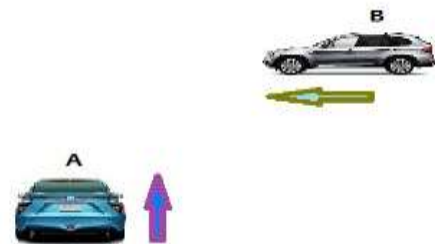
Fig. 7: One source one destination moving in different direction

In this case (Fig. 7), vehicle B is moving towards the west and the vehicle A is moving towards north. Both of the vehicles are moving at a speed of 100 km/h. Initially distance between them is 212.1320 meters. Vehicle B is the source node and vehicle A is the destination node. Vehicle B is monitoring its distance from vehicle A continuously and the moment it found that the distance monitored is less than or equal to the warning distance it will immediately send a warning textual message to the destination vehicle A. The vehicle A will take action according to the received textual message and gives a response to the sender.

3.3 Scenario Three

Fig. 8: One source two destinations are moving in different directions

In this case (Fig. 8), vehicle B is moving towards the west, the vehicle A is moving towards the north and the vehicle C is moving towards south. All the three vehicles are moving at speed of 100 km/h. Initially distance between vehicle B and vehicle A is 212.1320 meters, between vehicle B and vehicle C is also 212.1320 meters and between vehicle A and C is 300 meters. Vehicle B is the source node and vehicle A, vehicle C are the destination nodes. Vehicle B is monitoring its distance from vehicle A and vehicle C continuously and the moment it found that the distance monitored is less than or equal to the warning distance it will immediately send a warning textual message to the destination vehicle. The destination vehicle will take action according to the received textual message and gives a response to the sender.

3.4 Scenario Four

Fig. 9: One source one destination are moving in the same directions

In this case (Fig. 9), vehicle A and vehicle B both are moving in the same direction towards east. Vehicle A is moving at a speed of 100 km/h and the vehicle B is moving at a speed of 70km/h. Initially distance between them is 300 meters. Vehicle B is the source node and vehicle A is the destination node. Vehicle B is monitoring its distance from vehicle A continuously and the moment it found that the distance monitored is less than or equal to the warning distance it will immediately send a warning textual message to the destination vehicle A. The vehicle A will take action according to the received textual message and gives a response to the sender.

3.5 Simulation Configuration
Simulation environment is configured with the parameters are given in Table 1.

Table 1: VANET Simulation Parameters

Parameter	Value
Routing Protocol	VAODV
Channel Type	Wireless Channel
Number of Nodes	2, 3
Transport Protocol	UDP
Interface Queue Type	Queue/DropTail/PriQueue
Queue Length	50 Packet
Data Payload	512 Bytes/Packet
MAC Types	Mac/802_15_4
Mobility	Random way point
Transmission Range	250 meter
Speed	100 km/h, 70 km/h
Area of Simulation	8770 m X 2738 m
Time of Simulation	40 sec

3.6 Performance Measurement Metrics
To evaluate the model performance we have used four metrics these are

Average End-to-End delay

It is the delays due to buffering in the process of route discovery, queuing, delays during MAC level retransmission and the time taken during the propagation and transfer [19] [20]. It can be defined as

$$Avg\_End - to - End\_Delay = \frac{1}{T}\sum_{k=1}^{n}(r_k - s_k) \qquad Eq.2$$

Where T is the packets that are successfully received, **k** is the unique packet identifier, r_k is the time of the received unique id k packet and s_k is the time at which a packet with unique id k is sent. End-to-End delay is measured in milliseconds.

Normalized Routing Load

It is the number of routing packets transmitted per data packets delivered at the destinations [19] [20].

$$Routing\ Load = \frac{Routing\ Packets\ Sent}{Data\ Packets\ Received} \qquad Eq.3$$

Throughput

It is the number of bits received per unit time by the destination node. It is represented in kilo bits per seconds (kbps) [19] [20].

$$Throughput = \frac{Received\ Size}{End\ Time - Start\ Time} \times \frac{8}{1000} \qquad Eq.4$$

Where 'End Time-Start Time' is the data transmission period.

Packet Delivery Ratio

Ratio of the number of data packets successfully delivered to the destinations to those generated by the sources [19] [20].

$$Packet\ Delivery\ Ratio = \frac{Received\ Packets}{Sent\ Packets} \times 100\% \qquad Eq.5$$

Driver Reaction Time

Time of the accidental message received and the time of bakes are applied.

$$Driver\ Reaction\ Time = Message\ Receive\ Time + 0.75 \qquad Eq.\ 6$$

0.75 sec is the time taken by the driver to step his foot onto the brake pedal and press the brake [21].

4. SIMULATION RESULTS, ANALYSIS AND DISCUSSION
Simulation results are shown in Fig. 10-35 using gnu plot line graph
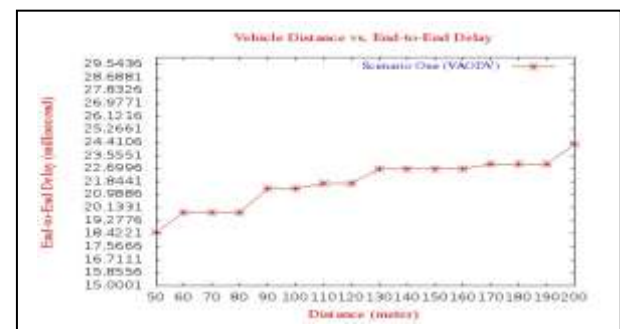
Fig. 10: Distance vs. End-to-End Delay for Scenario One
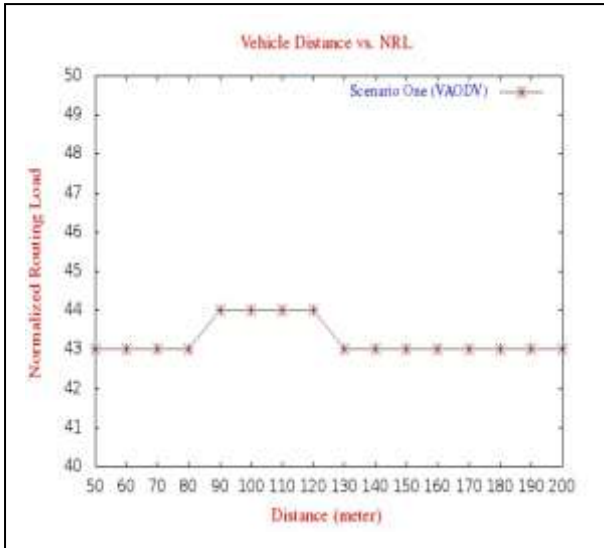
Fig. 11: Vehicle Distance vs. NRL for Scenario One
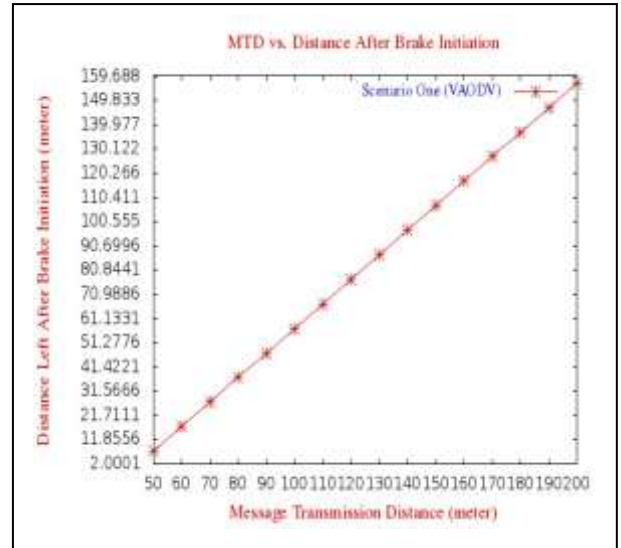
Fig. 14: MTD vs. Distance left for Scenario One
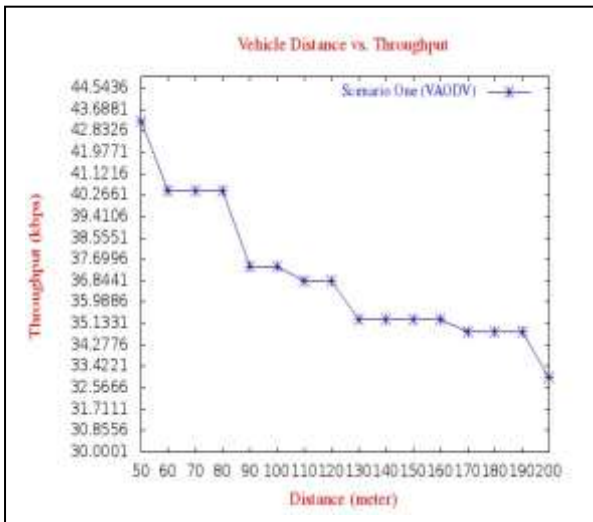
Fig. 12: Vehicle Distance vs. Throughput for Scenario One
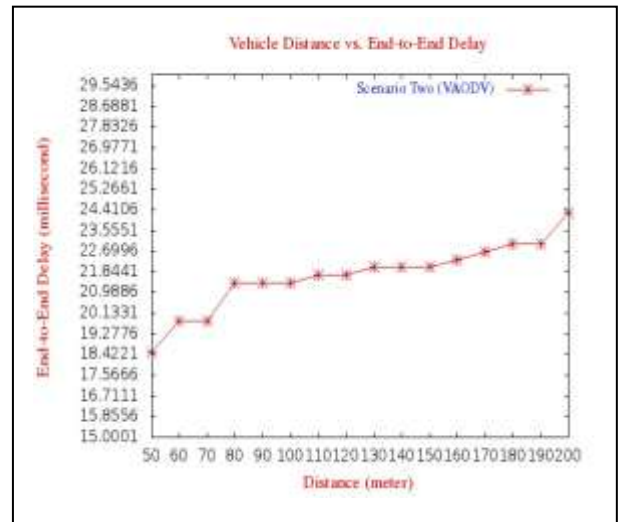
Fig. 15: Distance vs. End-to-End Delay for Scenario Two
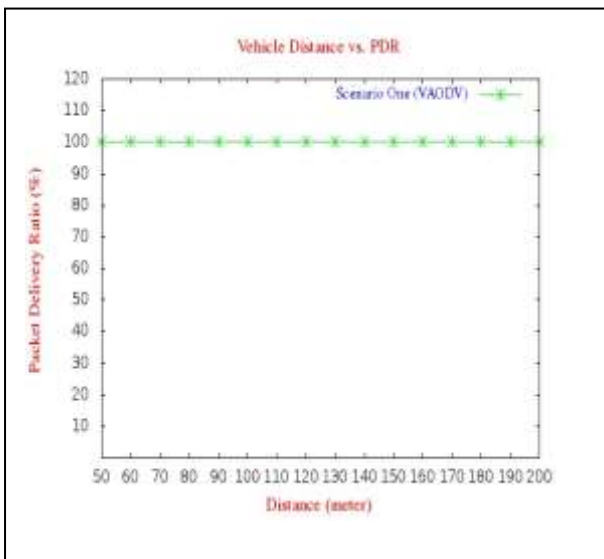
Fig. 13: Vehicle Distance vs. PDR for Scenario One
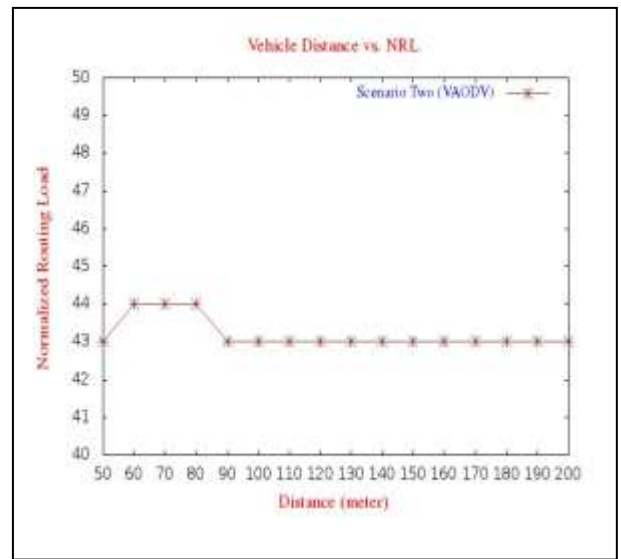
Fig. 16: Vehicle Distance vs. NRL for Scenario Two
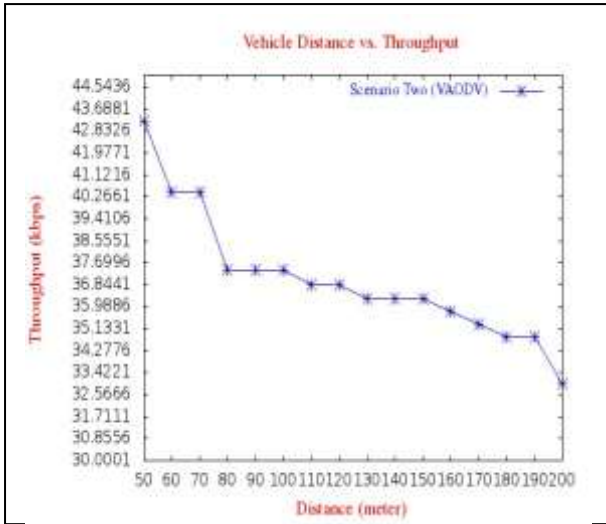
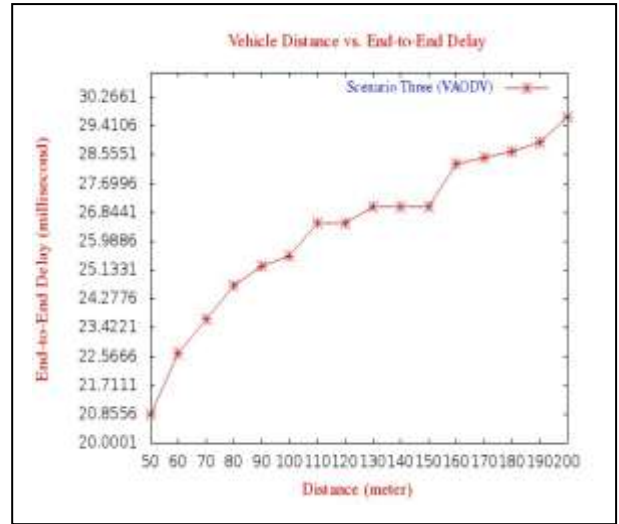Fig. 17: Vehicle Distance vs. Throughput for Scenario Two

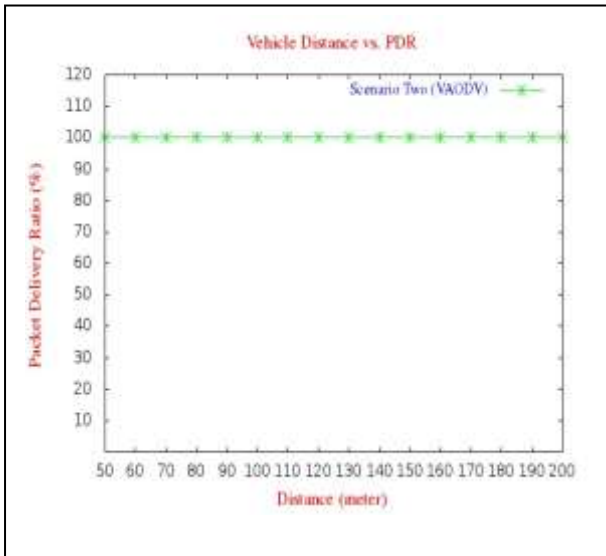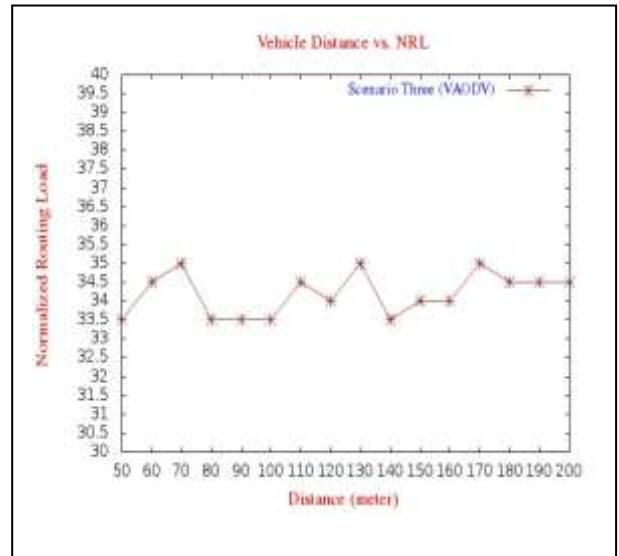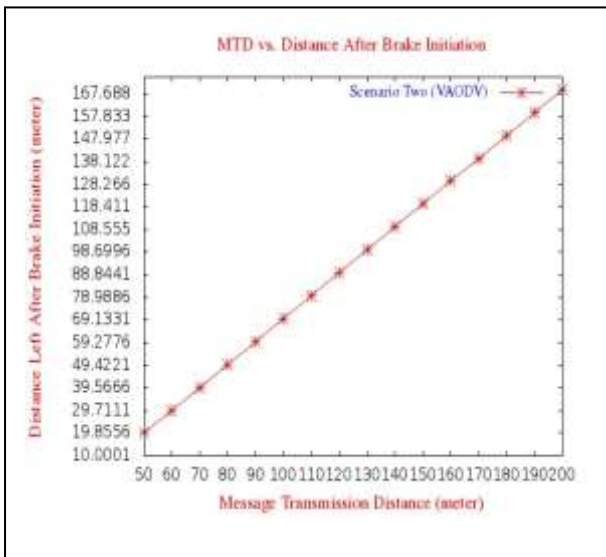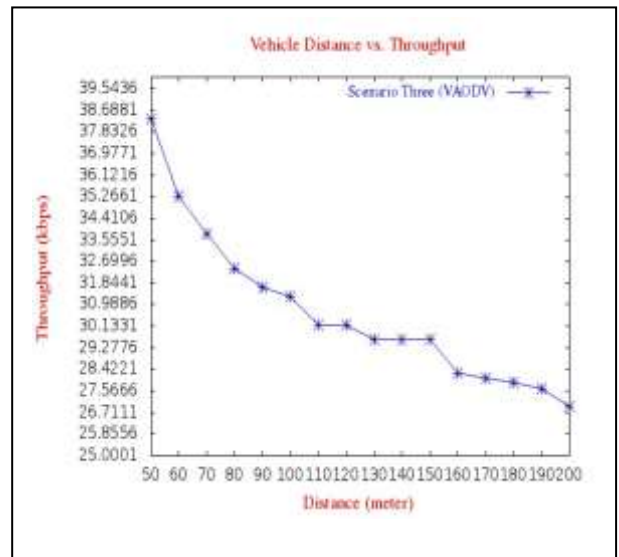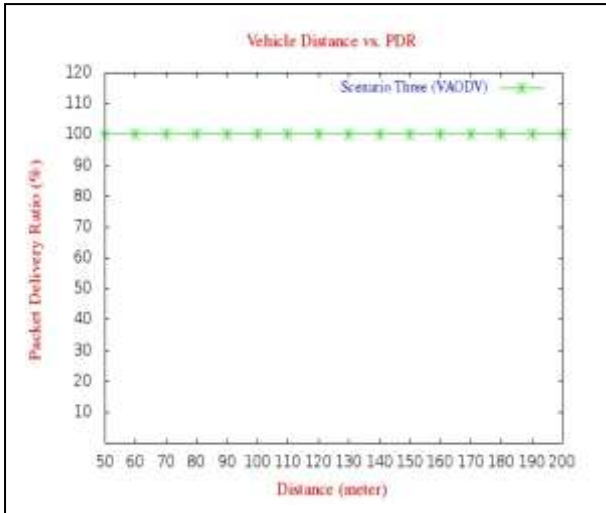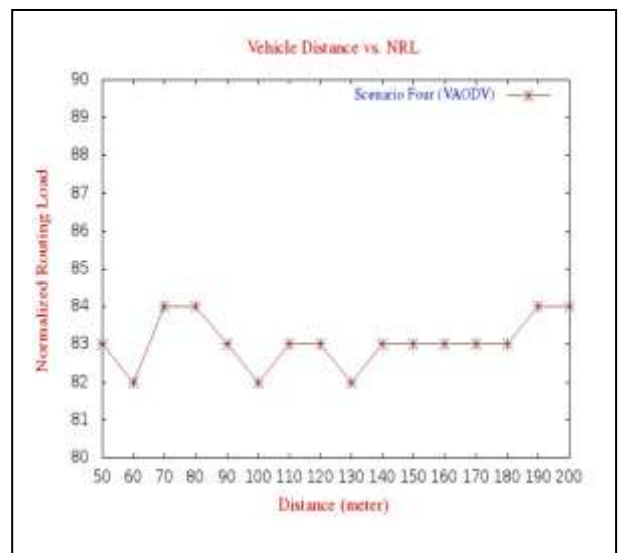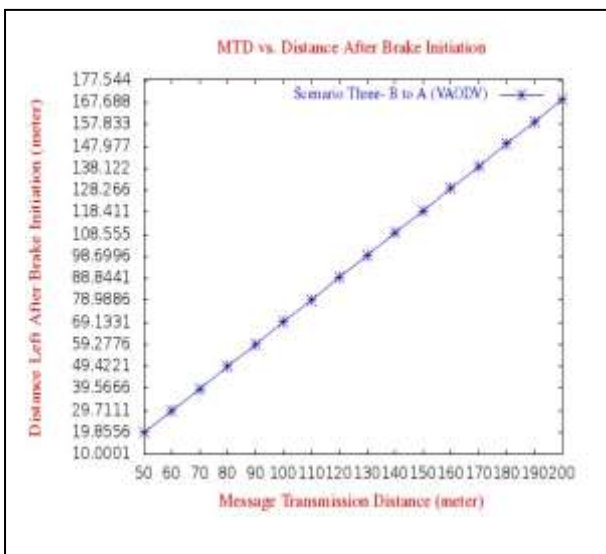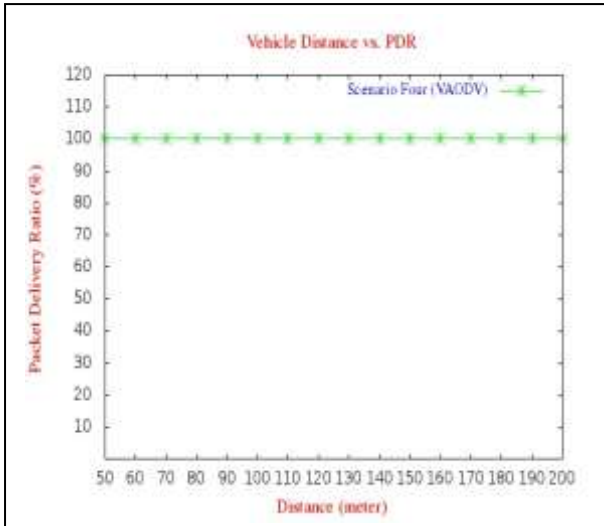

Fig. 20: Distance vs. End-to-End Delay for Scenario Three

Fig. 18: Vehicle Distance vs. PDR for Scenario Two

Fig. 21: Vehicle Distance vs. NRL for Scenario Three

Fig. 19: MTD vs. Distance left for Scenario Two

Fig. 22: Distance vs. Throughput for Scenario Three

Fig. 23: Vehicle Distance vs. PDR for Scenario Three
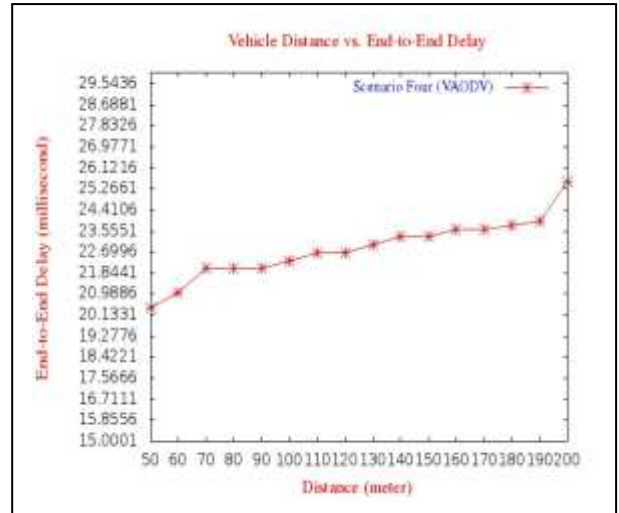
Fig. 26: Distance vs. End-to-End Delay for Scenario Four
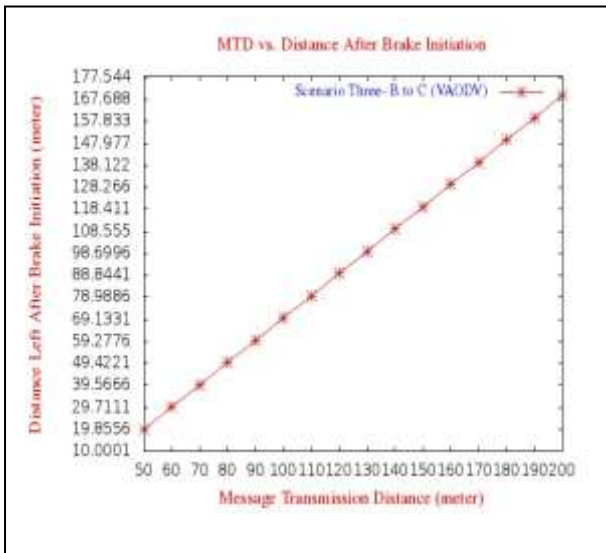
Fig. 24: MTD vs. Distance left in B to C for Scenario Three

Fig. 27: Vehicle Distance vs. NRL for Scenario Four

Fig. 25: MTD vs. Distance left in B to A for Scenario Three
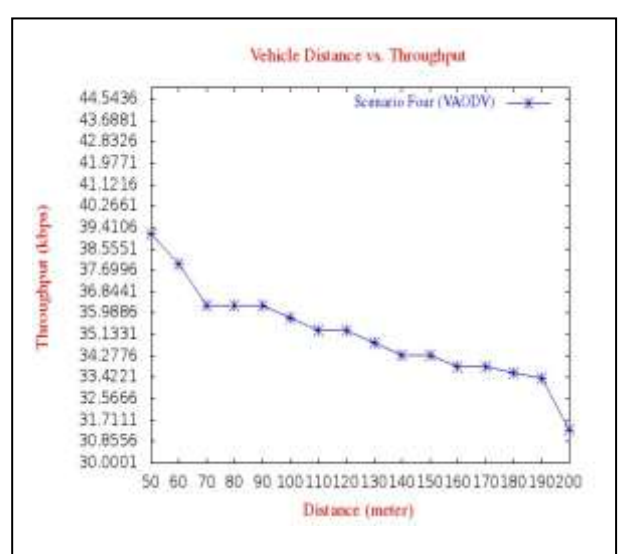
Fig. 28: Vehicle Distance vs. Throughput for Scenario Four

Fig. 29: Vehicle Distance vs. PDR for Scenario Four
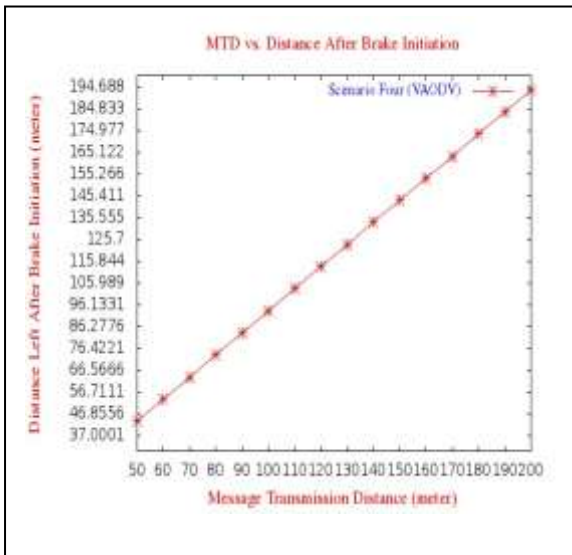
Fig. 30: MTD vs. Distance left for Scenario Four

In these figures Fig. 10-30 the End-to-End Delay, NRL, Throughput and PDR for four scenarios are calculated vs. vehicle distance means when the distance between any two vehicles is less than the warning distance the communication starts between these two vehicles and immediately the warning message sends. The warning distance is set as 50, 60, 70, . . . , 200 meters. The above figure shows the value of the End-to-End Delay, NRL, Throughput and PDR for different warning distance. The parameter Distance Left after Brake Initiation is also calculated vs. vehicle distance and shown in above, which is very important for to find out at what particular distance the warning message needs to send for what particular speed.

In the figures Fig. 14, 19, 25 and 30 of Driver Reaction Time, the message processing time to give an alarm is not included because it assumes that the OBU unit has immediately given an alarm to the driver of the vehicle when the warning message came and also the driver responded immediately.

Let us say, the message processing time to give an alarm to the driver of the vehicle when the warning message came and

driver thinking time is 1 sec. The results are shown in Fig. 31-35.

$$\text{Driver Reaction Time} = \text{MRT} + \text{AT} + \text{DTT} + 0.75 \qquad \text{Eq.7}$$

Where MRT is the Message Receive Time, AT is the time taken by the OBU to give an alarm and DTT is the Driver Thinking Time.
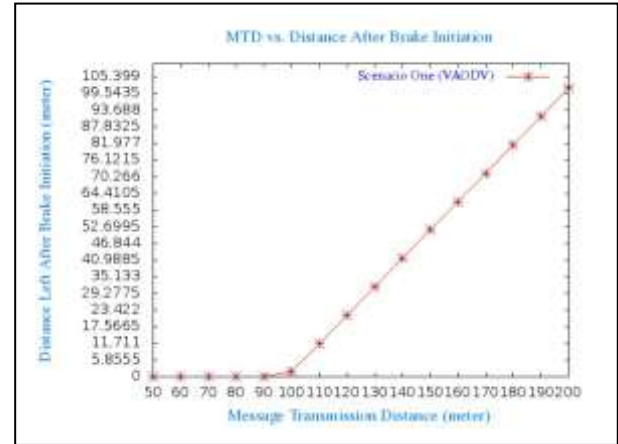
Fig. 31: MTD vs. Distance left for Scenario One
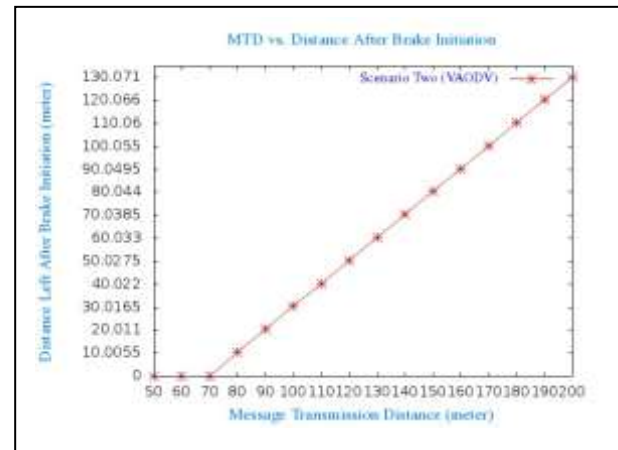
Fig. 32: MTD vs. Distance left for Scenario Two
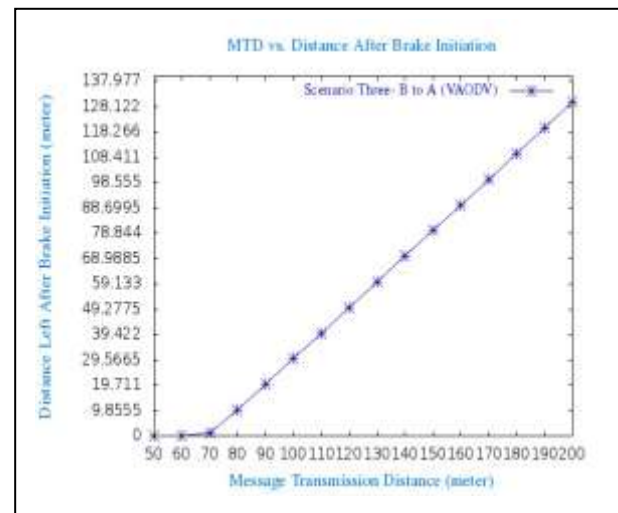
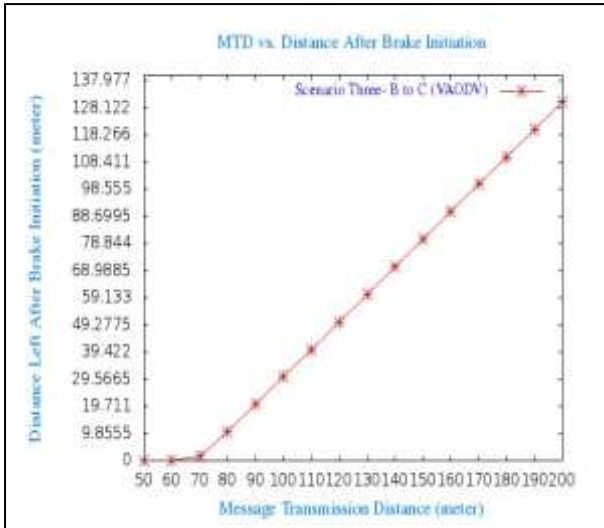Fig. 33: MTD vs. Distance left in B to A for Scenario Three

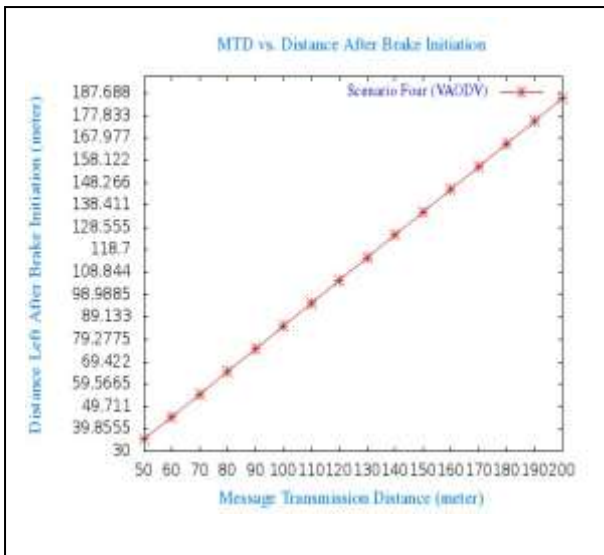Fig. 34: MTD vs. Distance left in B to C for Scenario Three

Fig. 35: MTD vs. Distance left for Scenario Four

The Table 2 contains the data of real world testing of vehicles [22].

Table 2: Braking Distance for speed of 100 km/h

Model	Distance (meter)
Alfa MITO	37.61
Alfa Giulietta QV	37.80
Audi A5 Sportback	37.62
BMW 123D Hatch	37.95
BMW 330D Coupe	36.63
Chrysler 300C	38.72
Holden VE Commodore SV6	39.86
HSV GXP	37.76
HSV GTS (WP tuned - 2011)	38.31
Nissan GTR (R35 - 2011)	32.75
Porsche 911 Turbo S (2011)	39.62
Renault Megane RS250	36.34
Renault RS Clio 200	36.43
Subaru Impreza WRX	37.38
Suzuki Alto	43.56
W Golf GTD	37.58
VW Golf R	39.57
VW Golf GTI	39.36
Volvo C30 TS	39.05

The results of real world testing data propose that braking distance less than 40 meters needs by a car at a speed of 100km/h.

The Table 3 contains the data given by the government [22].

Table 3: Braking Distance for speed of 0-180 km/h

Speed (km/h)	Distance (meter)
0	0
10	0
20	3
30	5
40	9
50	15
60	21
70	29
80	38
90	47
100	60
110	73
120	86
130	98
140	110
150	123
160	136
170	151
180	166

5. CONCLUSION

In this paper we have modified the AODV routing protocol named as VAODV which is used to simulate and also to analyze the VANETs. To find out the performance of the protocol the different parameters are calculated like End-to-End Delay, NRL, Throughput and PDR vs. vehicle distance for four different scenarios. Our simulation result shows that in case of Scenario One where vehicles are at a speed of 100km/h moving towards each other, the braking distance 40-60 meters (Fig. 14) will be left if the warning message is sent at a warning distance or vehicle distance of 90-110 meters. In case of Scenario Two where vehicles are moving at a speed of 100km/h, the braking distance 40-60 meters (Fig. 19) will be left if the warning message is sent at a warning distance or vehicle distance of 80-100 meters. In case of Scenario Three where vehicles are moving at a speed of 100 km/h, the braking distance 40-60 meters (Fig. 24, 25) will be left if the warning message is sent at a warning distance or vehicle distance of 80-100 meters. In case of Scenario Four where vehicles are moving in the same direction, one is at a speed of 100 km/h and another is at a speed of 70 km/h, the braking distance 40-60 meters (Fig. 30) will be left if the warning message is sent at a warning distance or vehicle distance of 50-70 meters.

6. FUTURE WORK

In the near future we plan to extend our work by giving more security to our VAODV protocol. We will also look into other factors like reliability by developing a trust based routing mechanism to track the misbehaving vehicles which is very important from the prospects of VANETs.

7. REFERENCES

[1] A. Iyer, S. S. Kulkarni, V. Mhatre and C. P. Rosenberg, 2008, "A Taxonomy-based Approach to Design of Large-scale Sensor Networks", Springer, pp. 3-33.

[2] T. AL-Khdour and U. Baroudi, 2011, "Literature Review of MAC, Routing and Cross Layer Design Protocols for WSN", InTech, pp. 1-27.

[3] I. F. Akyildiz and I. H. Kasimoglu, 2004, "Wireless sensor and actor networks: research challenges", ELSEVIER Ad Hoc Networks, vol. 2, no. 4, pp. 351-367.

[4] K. Sohraby, D. Minoli and T. Znati, 2007, "Wireless Sensor Networks: Technology, Protocols, and Applications",WILEY-INTERSCIENCE, vol. 1, pp. 1-307.

[5] J. Liu, Z. Yang and I. Stojmenovic, 2013, "Receiver Consensus: On-Time Warning Delivery for Vehicular Ad-Hoc Networks", IEEE Transactions on Emerging Topics in Computing, vol. 1, no. 1, pp. 57-68.

[6] K. A. Hafeez, L. Zhao, B. Ma and J. W. Mark, 2013, "Performance Analysis and Enhancement of the DSRC for VANET's Safety Applications", IEEE Transactions on Vehicular Technology, vol. 62, no. 7, pp. 3069-3083.

[7] H. A. Omar, W. Zhuang, A. Abdrabou and L. Li, 2013, "Performance Evaluation of VeMAC Supporting Safety Applications in Vehicular Networks", IEEE Transactions on Emerging Topics in Computing, vol. 1, no. 1, pp. 69-83.

[8] W. Ikram, S. Petersen, P. Orten and N. F. Thornhill, 2014, "Adaptive Multi-Channel Transmission Power Control for Industrial Wireless Instrumentation", IEEE Transactions on Industrial Informatics, vol. 10, no. 2, pp. 978-990.

[9] J. A. Gutierrez, M. Naeve, E. Callaway, M. Bourgeois, V. Mitter and B. Heile, 2001, "IEEE 802.15.4: a developing standard for low-power low-cost wireless personal area networks", IEEE Network, vol. 15, no. 5, pp. 12-19.

[10] Y. Huang, A. Pang and H. Hung, 2008, "An Adaptive GTS Allocation Scheme for IEEE 802.15.4", IEEE Transactions on Parallel and Distributed Systems, vol. 19, no. 5, pp. 641-651.

[11] Z. Shelby and C. Bormann, 2011, "6LoWPAN: The Wireless Embedded Internet", John Wiley & Sons, pp. 1-244.

[12] C. K. Singh, A. Kumar and P. M. Ameer, 2008, "Performance evaluation of an IEEE 802.15.4 sensor network with a star topology", Springer Science & Business Media, vol. 14, pp. 543-568.

[13] L. K. Bandyopadhyay, S. K. Chaulya and P. K. Mishra, 2009, "Wireless Communication in Underground Mines: RFID-based Sensor Networking", Springer Science & Business Media, pp. 1-477.

[14] E. M. Royer and C. K. Toh, 1999, "A review of current routing protocols for ad hoc mobile wireless networks", IEEE Personal Communications, vol. 6, no. 2, pp. 46-55.

[15] X. H. Lin, Y. K. Kwok and V. K. N. Lau, 2005, "A quantitative comparison of ad hoc routing protocols with and without channel adaptation", IEEE Transactions on Mobile Computing, vol. 4, no. 2, pp. 111-128.

[16] C. E. Perkins, E. M. Royer, S. R. Das and M. K. Marina, 2001, "Performance comparison of two on-demand routing protocols for ad hoc networks" , IEEE Personal Communications, vol. 8, no. 1, pp. 16-28.

[17] C. E. Perkins and E. M. Royer, 1999, "Ad-hoc on-demand distance vector routing", Second IEEE Workshop on Mobile Computing Systems and Applications, 1999. Proceedings. WMCSA '99, New Orleans, LA, pp. 90-100.

[18] J. H. Song, V. W. S. Wong and V. C. M. Leung, 2004, "Efficient on-demand routing for mobile ad hoc wireless access networks", IEEE Journal on Selected Areas in Communications, vol. 22, no. 7, pp. 1374-1383.

[19] T. Issariyakul and E. Hossain, 2011, "Introduction to Network Simulator NS2", Springer Science & Business Media, vol. 2, pp. 1-536.

[20] L. Shrivastava, S. S. Bhadauraia, G. S. Tomar and B. K. Chaurasia, 2013, "Secure Congestion Adaptive Routing Using Group Signature Scheme", Springer Transactions on Computational Science XVII, vol. 7420, pp. 101-115.

[21] M. Green, 2000, "How long does it take to stop? Methodological analysis of driver perception-brake times", Transportation Human Factors, vol. 2, no. 3, pp. 195-216.

[22] Safe Drive Training [Online]. Available: http://www.sdt.com.au/safedrive-directory-STOPPINGDISTANCE.htm.

A Review on Comparison of the Geographic Routing Protocols in MANET

Yukti
Department of Computer science and engineering
Giani Zail Singh- Punjab Technical University Campus
Bathinda , Punjab , India

Abhilasha
Department of Computer science and engineering
Giani Zail Singh- Punjab Technical University Campus
Bathinda , Punjab , India

Abstract: In Mobile ad-hoc networks (MANET) with high number of nodes and high mobility the routing of packets is a difficult task. In this paper, we are reviewing different geographic routing protocols as geographic routing are efficient for highly mobile nodes and made the communication scalable. Different protocols compared are The Distance Routing Effect Algorithm (DREAM), Location Aided Routing (LAR) Calculation, Greedy Perimeter Stateless Routing(GPSR) as of late new convention comes which is exceedingly proficient is the Adaptive position update (APU) strategy and further the improved APU strategy and on the basis of performance metrics the protocols are compared and reveals that the Improved APU strategy gives the high packet delivery ratio, lower delay and low energy consumption.

Keywords: DREAM, LAR, GPSR, APU, Improved APU

1. INTRODUCTION

MANET is a type of mobile ad hoc network that can change locations and configure itself on the fly. Because Mobile Ad hoc networks are mobile, they use wireless connections to connect to various networks. Some MANETs are restricted to local area of wireless devices, while others may be connected to the Internet. Numerous Directing conventions have been considered for geographic routing in MANET furthermore progressions are done over. Some of these conventions are The Distance Routing Effect Algorithm (DREAM), Location Aided Routing (LAR) Calculation, Greedy Perimeter Stateless Routing (GPSR) as of late new convention comes which is exceedingly proficient is the Adaptive position update (APU) system for geographic steering furthermore improved APU technique which more refines the consequence of APU methodology.

2. THE DISTANCE ROUTING EFFECT ALGORITHM (DREAM)

DREAM is location- based routing protocol, however not an absolutely geological one due to its proactive methodology. DREAM can be delegated as proactive routing strategy by using a new mechanism of dissemination and update of location information also routing tables for every one of the nodes in the network [3]. Every node proactively redesigns every other node about its location with the help of DREAM's location service. The overhead of such location updates is reduced in two ways. To start with, the effect of distance (nodes move slowly with respect to each other as their distance of separation increases). Next, every node generates updates about its location relying on its mobility rate fast moving nodes update more

frequently whereas slow moving nodes generate updates less routinely. DREAM geographically forwards data packets in the form of a directional flood. In DREAM the sender S of a packet with destination D will forward the packet to every one of the one-hop neighbors that lie "towards D." In order to determine this direction, anode calculates the region that is liable to contain D, called the *expected region*. As portrayed in Fig.1. The expected region is a circle around the position of D as it is known to S. since this position data may be out of date, the radius r of the expected region is set to $(t1-t0)$ $v$max, where $t1$ is the current time, $t0$ is the timestamp of the position information S has a session D, and $v$max is the maximum speed that a node may travel in the ad hoc network. Given the expected region, the "bearing towards D" for the delineation given in Fig. 1 is defined by the line in the middle of S and D and the angle α. The neighboring hops repeat this strategy utilizing their information on D's position. If a node does not have a one-hop neighbor in the oblidged direction, a recovery procedure must be started. This system is not some piece of the DREAM specific [1].
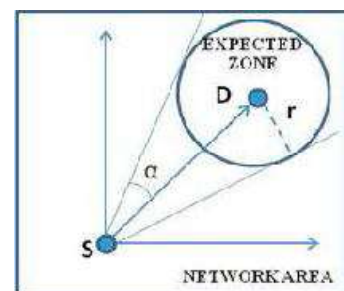

Figure1: Illustration of progress with DREAM [1].

3. LOCATION AIDED ROUTING (LAR) ALGORITHM

It makes presumptions of a 2-D plane, of GPS prepared nodes or the accessibility of another location service, of equal node range, of location error, of no congestion, no transmission error and no delays. Additionally, it is assumed only one sender and one destination. The aim is to reduce the number of nodes to which the route request is propagated. It is a routing protocol with two proposed methods: LAR1 and LAR2, both illustrated in Fig.2 [3].

In LAR1, the sending node advances the message only within the request zone and neighbors outside the region are not tended to. Within the limited sector, flooding is utilized.

In LAR2, the sending node always advances the message to all nodes closer to the destination than itself [1].
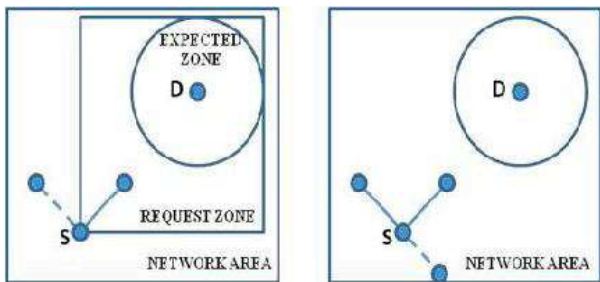
Figure2a.Progress with LAR1 b. Progress with LAR2 (forwarding option further away than source) [1].

4. GREEDY PERIMETER STATELESS ROUTING PROTOCOL (GPSR)

GPSR is a protocol that works in 2-dimensional plane. It is a Routing protocol which gives scalability with increasing nodes in the network. GPSR beacon telecasts MAC address with nodes IP and position [5]. It advances the information assuming the location service and keeps up a neighbor table, periodically updated through beacon messages. It results in a lots of data traffic; source's location is piggybacked on all data packets; it is validated in flat (2-D) topologies; it uses two methods for forwarding data: greedy forwarding and perimeter forwarding (right hand rule) [3].

In greedy forwarding, the forwarding is carried out on a greedy basis by selecting the node closest to the destination. This procedure proceeds until the destination is come to.

In perimeter forwarding, whenever the greedy forwarding method is not applicable or when this method fails, then the algorithm uses perimeter routing strategy to route around the communication voids and achieves the destination. Once the other node comes in transmission range, the algorithm changes back to the Greedy forwarding, reducing the delay and increment in the performance [8].

5. ADAPTIVE POSITION UPDATE (APU) STRATEGY

After GPSR protocol, new method APU strategy comes which incredibly simplifies the data transfer in MANET. There are some assumptions before as: all nodes are aware of their own position and velocity, all links are bidirectional, the beacon updates include the current location and velocity of the nodes, and data packets can piggyback position and velocity upgrades and all one-hop neighbors

operate in the promiscuous mode and hence can overhear the data packets [6].APU employs two mutually exclusive beacon triggering rules, which are discussed in the following:

5.1 Mobility Prediction Rule:

This rule adapts the beacon generation rate to the mobility of the nodes. Nodes that are highly mobile need to frequently update their neighbors since their locations are changing dynamically. Despite what might be expected, nodes which move slowly do not need to send frequent updates. In contrast periodic beacon update policy cannot satisfy both these requirement at the same time. In this scheme, upon receiving a beacon update from a node i, each of its neighbor's records node i's current position and velocity and periodically track node i's location using a simple prediction scheme based on linear kinematics (discussed below). Based on this position estimate, the neighbors can check whether node i is still inside of their transmission range and update their neighbor list accordingly. The objective of the MP rule is to send the next beacon update from node i when the error between the anticipated location in the neighbors of i and node i's actual location is greater than an acceptable threshold. We use simple location prediction scheme to estimate a node's current location. Note that, we assume that the nodes are located in a 2D coordinate system with the location indicated by the x and y coordinates. TABLE 1 illustrates the notations used.

Table 1
Notations for Mobility Prediction

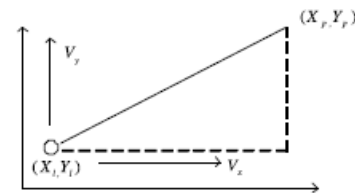| Variables	Definition
(X_l^i, Y_l^i)	The coordinate of node i at time T_l (included in the previous beacon)
(V_x^i, V_y^i)	The velocity of node i along the direction of the x and y axes at time T_l (included in the previous beacon)
T_l	The time of the last beacon broadcast.
T_c	The current time
(X_p^i, Y_p^i)	The predicted position of node i at the current time

Figure3. An example of mobility prediction

As shown in Fig. 1, the position of node i and its velocity along the x axis and y axis at time T_l, its neighbors can appraise the present position of i, using the following equations:

$$X_p^i = X_l^i + (T_e - T_l) * V_x^i$$

$$Y_p^i = Y_l^i + (T_e - T_l) * V_y^i$$

Note that, in this equation (X_l^i, Y_l^i) and (V_x^i, V_y^i) refers to the location and velocity information that was broadcast in the previous beacon from node i. Node i uses the same prediction criteria to keep record of its predicted location among its neighbors. Let (X_a, Y_a), denote the actual location of node i, acquired via GPS or other localization techniques. Node i then computes the deviation D_{devi}^i as follows:

$$D_{devi}^i = \sqrt{(X_a^i - X_p^i)^2 + (Y_a^i - Y_p^i)^2}$$

If the deviation calculated is greater than a certain threshold, known as the *Acceptable Error Range (AER)*, it acts as a trigger for node i to broadcast its present location and velocity as a new beacon. The MP rule then tries to maximize the effective duration of each beacon, by broadcasting a beacon only when the position information in the previous beacon becomes inaccurate. This extends the effective duration of the beacon for nodes with low mobility, thus reducing the number of beacons. Further, highly mobile nodes can broadcast frequent beacons to guarantee that their neighbors are mindful of the quickly changing topology.

5.2 On Demand Learning (ODL) Rule:

The MP rule solely may not be adequate for keeping up an exact local topology. Hence, it is important to devise a mechanism which will maintain a more exact local topology in those regions of the network where significant data forwarding activities are on-going. This is exactly what the *On-Demand Learning (ODL)* rule aims to accomplish. As the name recommends, a node broadcasts beacons *on-demand*, i.e. with respact to data forwarding activities that occur in the vicinity of that node. As indicated by this rule, whenever a node overhears a data transmission from a *new* neighbor, it broadcasts a beacon as a response. In actuality, a node waits for a small random time interval before responding with the beacon to prevent collisions with other beacons. Review that, it is assumed that the location updates are piggybacked on the data packets and that all nodes operate in the promiscuous mode, which permits them to overhear all data packets transmitted in their vicinity. Likewise, since the data packet contains the location of the final destination, any node that overhears a data packet also checks its current location and figures out if the destination is within its transmission range. Provided that this is true, the destination node is added to the list of neighboring nodes, if it is not already present. Note that, this specific check incurs zero expense, i.e. no beacons need to be transmitted. We allude to the neighbor list developed at a node by virtue of the initialization phase and the MP rule as the *basic* list. This list is mainly updated in response to the mobility of the node and its neighbors. The ODL rule permits active nodes that are included in data forwarding to enhance their local topology beyond this basic set. As it were, a *rich* neighbor list is maintained at the nodes located in the regions of high traffic load. Thus the rich list is maintained only at the active nodes and is built reactively in response to the network traffic. Every inactive node simply maintains the basic neighbor list. By making a rich neighbor list along the forwarding path, ODL guarantees that in situations where the nodes involved in data forwarding are highly mobile, alternate routes can be easily established without causing extra postpones.

Fig. 4(a) delineates the network topology before node A starts sending data to node P. The solid lines in the figure denote that both ends of the link are mindful of one another. The initial possible routing path from A to P is A-B-P. Presently, when source A sends data packets to B, both C and D receive the data packet from A. As A is a new neighbor of C and D, according to the ODL rule, both C and D will send back beacons to A. As a result, the links AC and AD will be discovered. Further, on the basis of location of the destination and their current locations, C and D discover that the destination P is within their one-hop neighborhood. Essentially when B advances the data packet to P, the links BC and BD are discovered. Fig. 4(b) reflects the enhanced topology along the routing path from A to P.
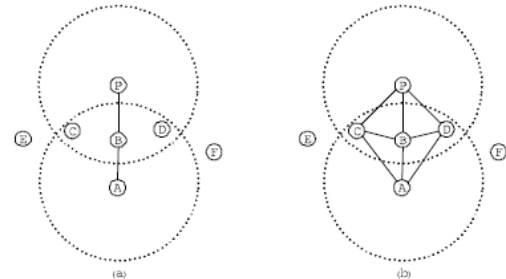

Figure4.An example illustrating the ODL rule [4].

Note that, however E and F receive the beacons from C and D, respectively, neither of them responds back with a beacon. Since E and F do not lie on the forwarding path, it is futile for them to send beacon updates in response to the broadcasts from C and D. Basically, ODL aims at improving the accuracy of topology along the routing path from the source to the destination, for each traffic flow within the network [4].

6. IMPROVED APU STRATEGY

The proposed Improved Adaptive Position Update (IAPU) strategy for geographical routing which progressively adjusts the regularity of position updates based on the mobility dynamics of the nodes and the forwarding patterns in the network. Improved APU is considering two essential standards such as Nodes whose movements are harder to predict update their positions more frequently, and Nodes closer to forwarding paths update their positions more frequently. The following are the systematic process of the Improved APU in which it bit by bit increases the performance of the existing Adaptive Position Update for Geographic routing with low mobility based forwarding node selection. This thusly further overcomes the link failure of the whole network in high mobility routing.

6.1 Beacon Updation

In this process, the nodes position changes either long or short each node should update their position more frequently through beacon packet. Updating each and every either low or high movement updating, it will consume more energy, and received by someone in general or increasing amounts over time.

6.2 Mobility Prediction

Mobility Prediction (MP) employs a basic mobility prediction scheme to estimate when the location information broadcast in the previous beacon becomes incorrect. The next beacon is send out only if the

predicted error in the location estimate is greater than a exact threshold, thus alter the update frequency to the dynamism inherent in the node's movement. A periodic beacon update policy cannot fulfill both these requirements at the same time, since a small update interval will be inefficient, whereas a larger update interval will lead to inaccurate position information for the highly mobile nodes. In our procedure, upon receiving a beacon update from a node i, each of its neighbor's records node is current position and velocity and periodically track node is location using a simple prediction scheme based on linear kinematics. Based on this position approximate the neighbors can check whether node i is still within their transmission range and update their neighbor list accordingly. The aim of the MP rule is to send the next beacon upgrade from node i when the error between the predicted location in the neighbors of i and node i's actual location is greater than an acceptable threshold.

6.3 On Demand Learning

Update forwarding path's closest neighbor position for effective routing performance improving the accuracy of the topology along the routing paths between the communicating nodes. ODL utilizes an on-demand learning approach, whereby a node broadcast beacons when it overhears the transmission of a data packet from a new neighbor in its neighborhood. This guarantees that nodes involved in forwarding data packets maintain a more up to date view of the local topology. Referred as On-Demand Learning (ODL), in which it aims at improving the exactness of the topology along the routing paths between the communicating nodes. On the opposing, nodes that are not in the vicinity of the forwarding path are unaffected by this rule and do not broadcast beacons very frequently.

6.4 Improved APU

In Mobile Ad-hoc Networks if forwarding nodes have high mobility, may have lot of chances to make local topology inaccuracy. To upgrade with low mobility based forwarding node selection we improve routing performance more than APU. If we take high mobility routing, link failure will affect the Whole Network. Through this way, we can able to send data without link failure. The Improved APU is that beacons generated in APU are more concentrated along the routing paths, while the beacons in all additional schemes are more scattered in the whole network. As a result, in modified APU, the nodes located in the hotspots are responsible for forwarding most of the data traffic in the network have an up-to-date view of their local topology [2].

7. PERFORMANCE METRICS

Diverse Performance metrics are compared for all the protocols. Some of the metrics are: packet delivery ratio, end to end delay, energy consumption.

7.1 Packet delivery ratio

Measures the percentage of data packets generated by nodes that are successfully delivered.

7.2 End to End Delay

This metric measure the average time it takes to route a data packet from the source node to the destination node.

7.3 Energy Consumption

The energy metric is taken as the average energy consumption per node calculated through simulation time.

8. SIMULATION RESULT

Comparison of geographic routing protocols in tabular form on the premise of qualitative parameters when the node density is high [7][5] [2]. The comparison depicts that as the protocols are derived for routing performance of routing increased.

Table 2
Comparison of geographic routing protocols

Parameter	Packet delivery ratio	End to end delay	Energy consumption
DREAM	Low	Long delay	High
LAR	Low	Long delay	High
GPSR	High	Lower delay	Low
APU	High	Lower than GPSR	Low
Improved APU	High	Lower delay than APU	Lower than APU

9. CONCLUSION AND FUTURE SCOPE

In this paper, we have clarified different geographic routing protocols and compared the performance of all and presumed that new procedure is superior to the previous one and more work can be done in this respect. In Future, further improvements in the current strategy and new techniques can be proposed and APU strategy can be further improved to give more reliable communication.

10. ACKNOWLEDGEMENT

11. REFERENCES

[1] Divya Wadhwa, Deepika, Vimm Kochher, Rajesh Kumar Tyagi (2014), "A Review of Comparison of Geographic Routing Protocols in Mobile Adhoc Network", Advance in Electronic and Electric Engineering, Vol. 4, No.1, pp. 51-58.

[2] Mr.C.Nallusamy, Dr.A.Sabari, Mr.N.Suresh (2014), "Improved Adaptive Position Update for Geographic Routing in Mobile Ad Hoc Networks", International Journal of Engineering Sciences & Research Technology, pp. 46-51.

[3] Ana Maria Popescu, Ion Gabriel Tudorache, Bo Peng, A.H. Kemp(2012),"Surveying Position Based Routing Protocols for Wireless Sensor and Ad-hoc Networks"*,* International Journal of Communication Networks and Information Security *(IJCNIS),* Vol. 4, No. 1**.**

[4] Quan jun Chen, Salil S. Kanhere, Mahbub Hassan(2013), "Adaptive Position Update for Geographic Routing in Mobile Ad Hoc Networks", IEEE TRANSACTIONS ON MOBILE COMPUTING, Vol. 12.

[5] Ms. Pranoti A. Bedekar, Dr. Y. M. Patil, Dr. A. S. Vaingankar (2014), "Review paper on Adaptive Position Update for efficient routing in mobile ad hoc networks", Journal of Emerging Technologies and Innovative Research, Vol.1 , Issue 6.

[6] Quan Jun Chen, Salil S. Kanhere, Mahbub Hassan, Kun-Chan Lan (2005), "Adaptive Position Update in Geographic Routing, School of Computer Science and Engineering", The University of New South Wales, Sydney, Australia.

[7] Gang Wang, Guodong Wang(2010), "An Energy Aware Geographic Routing Protocol for Mobile Ad Hoc Networks", Int J Software informatics*,* Vol. 4, No. 2, pp.183-196.

[8] Varun G Menon, Joe Prathap p m(2013), "Performance Analysis of geographic Routing Protocols in highly Mobile ad hoc Network", Journal of Theoretical and Applied Information Technology, Vol. 54, No. 1.

[9] Michele Zorzi, Senior Member, IEEE, and Ramesh R. Rao(2003), "Geographic Random Forwarding (GeRaF) for Ad Hoc and Sensor Networks: Energy and Latency Performance", IEEE TRANSACTIONS ON MOBILE COMPUTING, Vol. 2, No. 4.

[10] Y. Kim, R. Govindan, B. Karp and S. Shenker(2005), "Geographic Routing Made Practical", in Proceedings of the 2nd Conference on Symposium on Networked Systems Design and Implementation, Berkeley, CA, USA, pp. 217-230.

[11] B. Blum, T. He, S. Son, and J. Stankovic(2003), "IGF: A State-free Robust Communication Protocol for Wireless Sensor Networks", Tech. Rep., Department of Computer Science, University of Virginia, USA.

[12] B. Karp and H. T. Kung(2000), "GPSR: Greedy Perimeter Stateless Routing for Wireless Networks", in Proceedings of ACM Mobicom 2000, Boston, MA, USA, pp. 243-254.

[13] Ravi Kumar Poluru, T. Sunil Kumar Reddy, D.Nagaraju, "Updating the Routing Position in Ad hoc Networks by using Adaptive Position", International Journal of Computer Trends and Technology (IJCTT), Vol. 10, No. 3.

Developing Sales Information System Application using Prototyping Model

Muhammad Aziz, Andri Pranolo
Informatics Department
Universitas Ahmad Dahlan
Yogyakarta, Indonesia

Cita Agustin
Informatics Department
Universitas Teknologi Yogyakarta
Yogyakarta, Indonesia

Abstract: This research aimed to develop the system that be able to manage the sales transaction, so the transaction services will be more quickly and efficiently. The system has developed using prototyping model, which have steps including: 1) communication and initial data collection, 2) quick design, 3) formation of prototyping, 4) evaluation of prototyping, 5) repairing prototyping, and 6) the final step is producing devices properly so it can used by user. The prototyping model intended to adjust the system in accordance with its use later, made in stages so that the problems that arise will be immediately addressed. The results of this research is a software which have consumer transaction services including the purchasing services, sale, inventory management, and report for management needed purpose. Based on questionnaires given to 18 respondents obtained information on the evaluation system built, among others: 1) 88% strongly agree and 11% agree, the application can increase effectiveness and efficiently the organizations/enterprises; 2) 33% strongly agree, 62 agree, and 5% not agree, the application can meet the needs of organization/enterprise.

Keywords: transaction; transaction system; sales; prototyping model

1. INTRODUCTION

At this time, the recording process of purchasing and selling transactions is important to support the effectiveness and efficiently in services to consumer. Furthermore, many software has developed by organizations or enterprise to support many aspect including to increase sales, consumer satisfaction, other positive impact on profitability [1], which using various media, such as desktop personal computer, web-based, and also mobile application [2]. This research aimed to develop sales transaction information system using a prototyping model. The basics task based on functionality aspect, an information system can be identified meets [3]:

- Gathering information,
- Storing information,
- The processing of information,
- The transmission of information,
- Presenting information

Nowadays, Information systems exist because they are an integral part of a modern organization [4]. The information system is composed of humans, machines, and methods. A company running an operating activity relevant companies uses the data to generate information, information systems can connect a computer with machine storage media, office machines (fax, copy machine), communications equipment (controllers, routers), and other storage media [5]. Culture plays an important role in affecting software piracy, and individual behavior in general [6].

Some of the research that previously successfully created and have almost the same problems with this system, including research about problems of sales system that includes sales, sales reports and inventory reports [7]. In this work the agency rewrite of an existing file into the computer so making the report takes a long time at the end of the month, therefore the researchers designed a sales system that directly fill in the data transaction and sales reports every day in a system and then develop applications a sales system that will provide service in sales, especially in making monthly reports.

Other researcher has developed a web-based application for motorcycle sales to solve the problems of the existing sales system and purpose to increase income of selling [8], web-based automobile sales management system [9]. Web-based information for the purchase and sale also has developed before to simplify the consumers to obtain information about the goods and increase the efficiency and speed of consumers service [10].

Basically, such research has been able to solve the problems of selling, but fundamental differences with the system to be built there is the scope of sales which emphasizes the aspects.

2. METHOD

In the designing the system to be developed can use prototyping models. Prototyping is something is not complete, but something that should be evaluated and modified again. Any changes can happen when prototyping created to fulfill of user needs and at the same time allows developers to better understand the needs of users [11]. Prototyping is not a complex thing, but something that must devaluation and modified again. The steps of prototyping model, described as [11].

1. Communication and initial data collection. An analysis of user needs.
2. Quick design. A design in general to redevelop.
3. Formation of prototyping, the manufacture of the device prototyping testing and refinement.
4. Evaluation of prototyping. Evaluating and refine the analysis of the program were needed.
5. Repair prototyping. The manufacture of the actual types based on results of the prototyping evaluation.
6. The final step is producing devices properly so it can use by user.

Implementation of an application to the database in the development of this system used Entity Relationship (ER) data model and Relational data model. ER diagram express entire logical structure of a graphical database. ER diagram has a quality that is clear, simple and allows to explain most of the database structure or the extension of the use of the ER model [12]. ER data model must be reduced to Relational data mode if it will be implemented to database management systems.

3. RESULTS AND DISCUSSION

Analysis of systems has important function in the details and design of a new information system. This analysis process is a step in understanding problems before taking an action or decision of completion the main results. Next on the stage of system design purpose to provide an overview of the system being designed. Illustration interface design of applications that are built as in the figure 1.
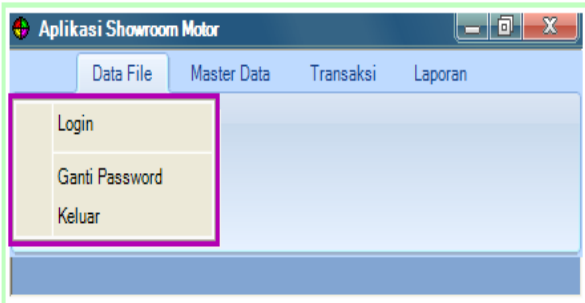

Figure 1. Design of program's main form

To design a database used the Entity Relationship (ER) Model and Relational Model. With ER model describes the real world data model ER model is very useful in mapping the meaning and the company's interaction to a conceptual schema so many database design tool draws on the concepts from the ER model [12]. Model E-R data using three basic concepts:

1. Entity Sets (ES). ES is objects/thing in the real world that can be distinguished from all other objects/thing.
2. Relationship Sets (RS). Several entities can be associated each others.
3. Attributes are properties that having by Entity Sets. Each attribute have domain to describe a set of values that allowed.

Results of the ERD will contain several Entity sets that describe the relationship among others (Figure 2). The rectangle is divided into two parts to represent entity. To design the ERD follow to Sileberschatz et. al. [12], in this case we used several object, such as: 1) The rectangle contains the name of the entity set and all the attributes of the entity set, 2) diamonds to represent relationship, 3) a line used for connected among entities and relationships.
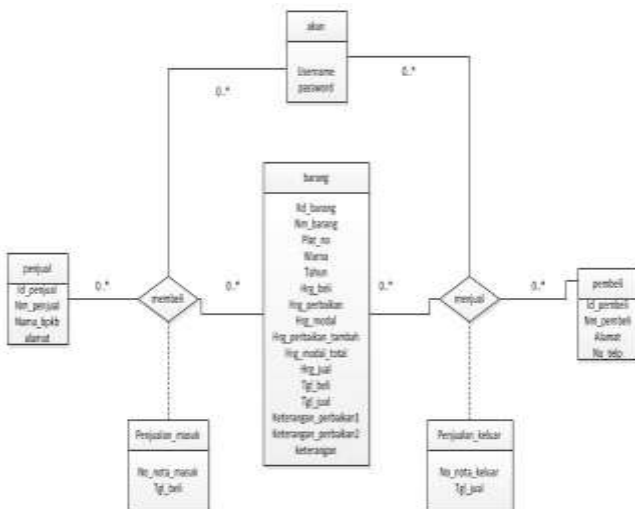

Figure 2. ERD sales transaction systems

Figure 3 shown a Relational data model which resulting from ERD (Figure 2), called database schematic diagram which describe tables in a database and the relationship among others. Relational data model suitable to be implemented in database management system (DBMS).

The database is a collection of data that contains information [12], it appears to be computerized data management to allow users to manipulate information that stored to supply the organization needed.

Figure 3. Schematic design database

The process for each process and data flow at the system is used Data Flow Diagram (DFD). DFD is a graphical representation which describes the flow of information and the transformations are applied as data moves from input to output. Data flow diagrams can be used to represent a system or software at any level of abstraction. DFD provide an additional of information used as long as domain analysis of information and serves as a basis for modeling of functions [11].
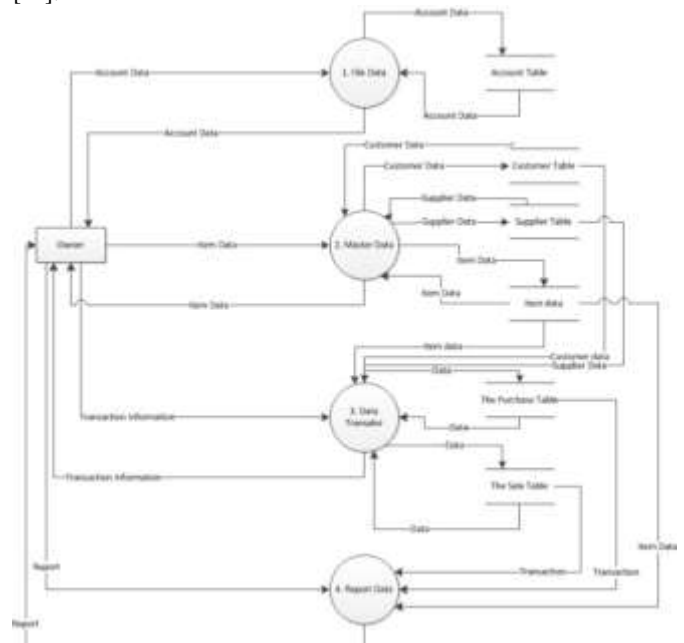

Figure 4. DFD level 1

In this research we created several level DFD including : DFD level 0 or context diagram, DFD level 1, DFD level 2 process 1 for file data, DFD level 2 process 2 for master data, DFD level 2 process 3 for transactions process, and DFD level 2 process 4 for reports process. The DFD Level 1 shown in Figure 4, which it describes the all system process developed.

In the relational database SQL commands (Structure Query Language) which includes DDL (Data Definition Language) and DML (Data Manipulation Language) [12]. DDL associated with specification of the structure of the database including the relation schemas, domains, constraints (constraints),etc. Example of DDL about account table (Figure 5).

```
CREATE TABLE 'akun'('username'
varchar(50) NOT NULL, ON PRIMARY KEY
'password' varchar(50) NULL
```

Figure 5. DDL CREATE TABLE

Next DML is used to query information includes select, add, update, and delete. Example DML in Borland Delphi applications for account table can view in figure 6.

```
Sql.Text:='SELECT * from akun where
username='+QuotedStr(Edit1.Text)+'and
password='+QuotedStr(Edit2.Text);
```

Figure 6. DML for Account Table

In this interface, we can see a button entry, which when the account data is entered, the data is checked in the database, if it is found then the user can access the program, but if it is not found, it will display a warning message that the account does not exist, Interface design for information systems account is presented in Figure 7.

Figure 7. Account Form

The system information sale, overall include some form application. Main Form in the figure 8. In the Main form has menu from file data, master data, transaction, and report.
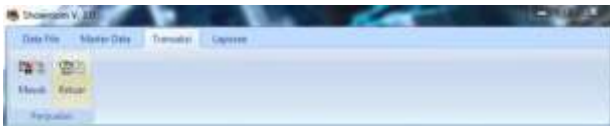

Figure 8. Main menu

In the form of purchase will be seen data that must be filled, the data that has been filled will be stored in the database. In the form of purchase, there are two main buttons is save button and change, which if it was wrong data entry

will still be replaced, but the date and number of nota could never be modified, is used as a reference if there is a change so the purchase history can still be seen.

Interface design for transaction system has 2 form, purchase and sale. The Form purchase can view in figure 9.
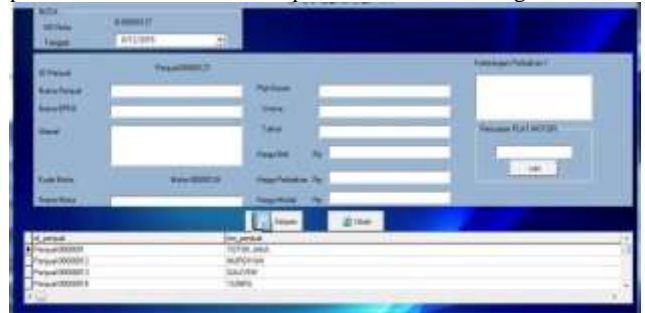

Figure 9. Form Purchase

Form of selling a form to enter sales data, which when a sale is the first thing to do is find code of product to be sold, then you will see the data which cannot be edited, but can still fill in the data of people which will buy. To direct nota will be automatically filled. In the search, which has sold product code cannot be done again because the sales transaction data in the database automatically updates the product data in the master data. Selling Form can be seen in the figure 10.

Figure 10. Selling Form

The sales transaction system will direct can be used to print a report purchases, sales, and inventory for each month. Furthermore, the transaction and reporting services will be faster. The sales report shon in Figure 11, the date of report can be adjusted to determine period of report. The system will be looking for the data in the database accordance with that date perform transactions on period date.
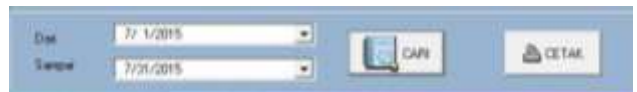

Figure 11. Sales report form

Based on questionnaires given to 18 respondents obtained information on the evaluation system built, among others: 1) 88% strongly agree and 11% agree, the application can increase effectiveness and efficiently the organizations/enterprises; 2) 33% strongly agree, 62 agree, and 5% not agree, the application can meet the needs of organization/enterprise.

4. CONCLUSION

The results of this works including:

1. An information system that can be used to purchases report, sales, inventory information, and reporting information based on management needed.
2. Entity Relationships data model produces four entities, and in Relational data model produces seven tables that have relationship among others, and one table not have relationships.
3. The process can be serve by the system are: 1) account management; 2) master data of items, costumer, and producer; 3) Purchase and sales transaction; and 4) report transactions.
4. Based on questionnaires given to 18 respondents obtained information on the evaluation system built, among others: 1) 88% strongly agree and 11% agree, the application can increase effectiveness and efficiently the organizations/enterprises; 2) 33% strongly agree, 62 agree, and 5% not agree, the application can meet the needs of organization/enterprise.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] S. Mithas, A. R. Tafti, I. Bardhan, and J. M. Goh, "Information technology and firm profitability: mechanisms and empirical evidence," *Mis Q.*, vol. 36, no. 1, pp. 205–224, 2012.

[2] T. Page, "Use of Mobile Device Apps in Product Design," *Int. J. Green Comput.*, vol. 4, no. 1, pp. 18–34, 2013.

[3] P. Bajdor and I. Grabara, "The Role of Information System Flows in Fulfilling Customers� Individual Orders," *J. Stud. Soc. Sci.*, vol. 7, no. 2, 2014.

[4] M. Romney, P. Steinbart, J. Mula, R. McNamara, and T. Tonkin, *Accounting Information Systems Australasian Edition*. Pearson Higher Education AU, 2012.

[5] I. Isa, "Evaluasi Pengontrolan Sistem Informasi," 2012.

[6] A. J. Setterstrom, J. M. Pearson, and H. Aleassa, "An exploratory examination of antecedents to software piracy: a cross-cultural comparison," in *System Science (HICSS), 2012 45th Hawaii International Conference on*, 2012, pp. 5083–5092.

[7] M. Fathoni, "Pembangunan Sistem Informasi Penjualan Sepeda Motor Bekas Pada Unit Dagang (Ud) Merdeka Motor Pacitan," *EPUB-SISTEM Inf.*, vol. 1, no. 1, 2013.

[8] S. D. Lasmana and F. R. Purba, "Sistem Informasi Penjualan Motor Berbasis Web pada Showroom Tunggal Jaya Motorindo," *Tek. dan Ilmu Komput.*, vol. 2, no. 8, 2013.

[9] H. Xiao, "Web-based Automobile Sales Management System," 2015.

[10] M. Minarni and B. Antoni, "PERANCANGAN APLIKASI PENJUALAN MOBIL BEKAS DI CV. SHANDY MOTOR BERBASIS WEB," *J. TeknoIf ISSN 2338-2724*, vol. 1, no. 2, 2014.

[11] R. S. Pressman, *Software engineering: a practitioner's approach*. Palgrave Macmillan, 2005.

[12] A. Silberschatz, H. F. Korth, and S. Sudarshan, *Database system concepts*, 6th ed. New York, USA: The McGraw-Hill Companies, Inc, 2011.