International Journal of Computer Applications Technology and Research
Volume 6–Issue 3, 121-129, 2017, ISSN:-2319–8656

ICT: A Cornerstone for Effective Weather Forecasting

Onu Fergus U.
Department of Computer Science, Ebonyi State University,
Abakaliki-Nigeria

Nwagbo Chioma L.
Department of Computer Science,
Nwafor Orizu College of Education,
Nsugbe-Nigeria

Abstract: Weather conditions continue to change at uncontrolable rate. The continous change bring about unberable hazard which needs high precision tools for its control. The global community have understood that the potential magnitude, harshness and impact of the symptoms of weather changes such as: global warming, loss of crops through over extensive periods of drought, unpredictable rainfall patterns, melting glaciers, displaced populations in search of refuge after floods, or entire villages devastated by the cruel force of cyclones and hurricanes on upcoming generations will be very intense. It is therefore assumed that information and communication technology is a key technology tools which can be used to ease, adapt to and monitor Weather condition. This paper therefore examines the linkage between Information and Communications Technologies (ICTs) and Weather Forecasting; it also looked into the signs and effects of Weather condition and the ICT solutions. It was deduced that ICTs not only help in advanced weather forecasting and climate monitoring, but are very essential in information circulation to a large audience through mobile devices. Finally, this paper, recommends solutions on how proposed information and communication technology strategies can be adopted to alleviate, adapt to and monitor weather condition.

Keywords: ICT, Weather forecasting, Mitigating Weather change, monitoring Weather condition, adapting to Weather condition, role of ICT in Weather forecasting,

1. INTRODUCTION

The world today is constantly experiencing changes in weather conditions, which its negative impacts include: global warming, floods, droughts, heat waves, etc. These harsh weather conditions affect both developing and developed countries although the poor/developing countries are most hit. [1] asserts that "Africa is hardest hit due to it's rich industrialized countries in the North that are responsible for 75 percent of green house gas emissions. Zimbabwe is suffering more in Africa, showing the effect of weather change risks which is felt more by the poor now and in the near future." Furthermore, some nations who do not contribute negatively to weather change and others whose contributions to weather change are minimal are also being affected by the negative consequences of weather change are frequently and greatly affected more than the real culprits.

It is a sad fact that more people are dying now than ever before and more diseases are surfacing as a result of weather change and more property is being destroyed now than ever before due to weather conditions. Weather change has and continues to be a major economic setback for most people especially in developing countries. The Fourth Assessment Report of the Inter-governmental Panel on Climate Change indicated that Africa will suffer the most from the impacts of climate change.

Forecasting is used to analyze long/short and isolated series of weather problems, whose data are often collected via files, input/sensing devices and processed or run using Hybrid computers.

When numerous items such as weather condition must be forecasted on a frequent regular basis, the uses of pre-defined or automatic forecasting techniques are critical. Weather forecasting systems must not only meet global information reports for accuracy but also for processing speed and robustness in adapting to and monitor Weather conditions. [2]

Weather Forecasting is therefore the prediction of change in weather condition to determine the signs, possible impact, location and effects of Weather change. These forecast of weather changes are achieved through Information and Communication Technologies (ICT) which is a key technology for weather forecasting. ICTs not only help in advanced weather forecasting and climate monitoring, but are very essential in information circulating to large global audiences. It can help address major adaptation risks such as food and water shortages through providing early warning systems and better monitoring of soil conditions and water quality.

2. THE OBJECTIVE OF THE STUDY IS TO:

- Determine the link between ICT and weather forecasting.
- Explain the use of ICTs information dispersal to people who could be affected in order to adapt to the effects of weather change.
- Document signs and effects of weather change.
- Articulate the causes of weather change.
- Establish the role of ICTs in adapting to weather change.
- Determine the role of ICTs in mitigating weather change.
- Reduction of high risk in weather changes through Risk analysis.

2.1 Use Of Icts For The Dispersal Of Weather Change Information

From 1980-2005, over 7,000 natural disasters happened worldwide in which millions of lives were lost. Ninety percent of these disasters were caused by weather and water connected events such as floods, cyclones and droughts [3]. on the other hand, we will show how the science of

weather forecasting and weather monitoring, which is serious to reducing such high casualty rates, is being advanced by the development in ICTs.

The Nairobi Framework[16], adopted in 2006, aims to assist all United Nations Framework Convention on Climate Change (UNFCCC) reporting Parties, in meticulous developing countries, including the Least Developed Countries (LDCs) and Small Island Developing States (SIDS), to pick up their understanding and assessment of impacts, openness and adaptation, and to make informed result on practical adaptation actions and measures to respond to weather change on a sound scientific, technical and socioeconomic basis, taking into report current and future change in weather condition and variability. As part of its commitment to contribute to the battle against weather change, the International Telecommunication Union (ITU) is a partner of the Nairobi Framework programme.

The ITU is the particular agency of the United Nations responsible for information and communications technologies (ICTs). Its membership, comprising 192 governments and over 700 private companies, has called for the ITU to take the lead in appealing the global community (including the UN system and the ICT industry, as well as academia and NGOs) to address weather change through the use of ICTs.[4],[5].

2.2 Signs and Effects Of Weather Change

Weather change can have serious developmental effects that hit particularly hard those countries that are already experiencing the hardships of poverty and marginalization. From the foregoing statement, it is crystal clear that weather change has developmental effects especially on poor developing nations. Regarding the main characteristics of weather change, the United Nations Framework Convention on Climate Change [6] , increases in average global temperature (global warming); changes in cloud cover and precipitation particularly over land; melting of ice caps and glaciers and reduced snow cover; and increases in ocean temperatures and ocean acidity – due to seawater absorbing heat and carbon dioxide from the atmosphere. Global warming has changed the type, frequency and intensity of extreme events, such as heat waves, tropical cyclones (including hurricanes and typhoons), floods, droughts and heavy precipitation.

UNFCCC establishes that Africa is already a continent under pressure from climate stresses and is highly vulnerable to the impacts of climate change. They noted famine and widespread disruption of socio-economic well-being as some of these impacts in Africa [2].

Migration of the malaria mosquito to higher altitudes will expose large numbers of previously unexposed people to infection in the densely populated east African highlands. Weather change is an added stress to already threatened habitats, ecosystems and species in Africa, and is likely to trigger species migration and lead to habitat reduction.

Report from the investigate carried out in Zimbabwe on effects of weather change on agriculture[7]. In their study, it was found out that weather change negatively affects maize production and some areas in Zimbabwe might be turned into non-maize producing as a result of drought. Further to that, Zimbabwe's Initial National Communication under the United Nations Framework Convention on Climate/weather Change indicates that weather change has adverse effects on human health for instance, in Zimbabwe; the incidence of malaria was very high after heavy rains and high temperatures throughout the country. Due to weather change, Zimbabwe is experiencing an increase in the frequency of floods.[8],[9].

Looking at the African continent the main impact of weather/ climate change on the continent will be increased frequency of natural disasters, droughts, floods and other weather extremes that lead to loss of life, economic

disruptions, social unrest and forced migration as well as major environmental problems.[10] Russell went on to say "prolonged drought periods will cause stress on water resources and reduce food security due to diminished agricultural productivity, increase outbreaks of vector borne diseases and other health impacts.

All the above cases of the effects and signs of weather change point to the fact that developing countries are experiencing the negative impacts of weather change. This problem is not affecting only developing countries but all countries in the world. Weather change is a global problem which requires collective solutions and collective effort from all countries [9].

2.3. Causes Of Weather Change

There are two classes of causes of climate change namely natural causes and man-made causes. International Telecommunication Union (ITU) went to say that natural causes include variations in solar radiation, volcanic activity and so on[5]. They Further, asserts that man-made weather change is of major concern because it appears to be leading to a progressive and accelerating warming of the planet, as a result of the release of greenhouse gases (GHG), primarily carbon-based emissions. On the other hand, the United Nations Framework Convention on Climate Change names two major causes of weather change which are rising fossil fuel burning and land use changes. The United Nations Framework Convention on Climate Change notes that these two activities have emitted, and are continuing to emit, increasing quantities of greenhouse gases (such as carbon dioxide, methane, and nitrogen dioxide) into the Earth's atmosphere. A rise in these gases has caused a rise in the amount of heat from the sun withheld in the Earth's atmosphere, heat that would normally be radiated back into space. This increase in heat has led to the greenhouse effect, resulting in weather/climate change [11],[12].

it is clear that human activity is the major contributor to weather change.

2.4. The Role of ICTs In Adapting To Weather Change.

Some adaptation strategies have been proposed in earlier researches such as Zimbabwe's Initial National Communication under the United Nations Framework Convention on Climate Change [7] ,[13]. However, the strategies proposed in the context of a developing country

Wireless communications and Early Warning Systems are adopted to facilitate efficient disaster warning and emergency response. Global e-Sustainability Initiative proposes solutions which are not very much special from the above solutions. Some of the examples of ICT solutions for adaptation proposed by Global e-Sustainability Initiative consist of, early warning systems where ICT systems are used to provide people with warnings and information regarding threats like extreme weather events, smart planning in which ICT systems are used to improve urban and rural planning and solutions can be tested based upon their resilience, e-health where smart ICT can bring down costs for health and permit correct treatment, especially in case of pandemics and new health challenges where local knowledge might be lacking and lastly education where ICT can help reduce costs, enable access and develop the quality of education through telecentre, use of e-learning systems.[14],[15].

IPCC report on weather change tackles the priorities of developing countries through adaptation (i.e. recovery and adjustment in the face of change in weather conduction)[11]. IPCC further asserts that the potential of ICTs in adapting to weather change is now evident in use of devices like mobile phones and other applications used in adapting to climate change[11] investigate.

2.5 The Role of ICTs In Mitigating Weather Changes

Information and communication technologies (including radio and telecommunication technologies, standards and supporting

publications) are being used for weather forecasting, climate monitoring, predicting, detecting and mitigating the effects of natural disasters. [12]ITU cites technologies which allow remote monitoring and data collection using ICT-equipped sensors (telemetry). In addition to that, they also cites aerial photography, satellite imagery, grid technology and in particular the use of global positioning by satellite (GPS) for tracking slow, long-term movement, for instance of glaciers or ice floes.[2].

Satellites and weather radars also track the advancement of hurricanes and typhoons and tracking tornadoes, thunderstorms, and the effluent from volcanoes and major forest fires. Also, the radio-based meteorological aid systems accumulate and process weather data. It is further asserted that distant from monitoring the effects of climate change, ICTs have as well proved invaluable in computer modeling of the earth's atmosphere [12].

Further to that supercomputers are being used in meteorological services to produce complex general circulation models of weather. Further, different radio-communication systems (satellite and terrestrial) are used for dissemination of information concerning diverse natural and man-made disasters [12].

In addition to the foregoing solutions, ITU indicates that the role of ICTs in weather /climate monitoring is shown in the structure of the World Meteorological Organization's (WMO) World Weather Watch (WWW), which involves three integrated core system components as follows:

- The Global Observing System (GOS) that provides observations of the atmosphere and the Earth's surface (including oceans) from the globe and from outer space. The GOS uses remote sensing equipment placed on satellites, aircraft, radios and relay data to environment control centers.
- The Global Telecommunication System (GTS) — radio and telecommunication networks for real-time exchange of a huge volume of data between meteorological centers.
- The Global Data Processing System (GDPS) — thousands of linked mini, micro and supercomputers, processes an enormous number of meteorological data and generates warnings and forecasts. [7], [10].

From the above discussion, it is clear that ICTs play a major role in monitoring weather change through helping with data collection, dissemination, storage, collaboration, processing and management.

2.6 Risk Analysis

ICT as well as remote sensing and geographic information systems have extended the possibilities for risk assessment of multiple hazards and enabled the development of different scenarios and contingency plans. Risk analysis includes: risk maps, hazard maps, and scenario maps. Risk analysis is as a result a key component in developing a disaster risk reduction strategy by establishing the links between exposure to hazards, level of vulnerabilities and the capacity to cope [9]. We all know that information and communications technologies (ICTs) have revolutionized our world...ICTs are also very important to confronting the troubles we face as a planet: the threat of climate change...Indeed ICTs are element of the solution. Already these technologies are being used to cut emissions and help countries adapt to the effects of climate change...Governments and industries that embrace a tactic of green growth will be environmental champions and economic leaders in the twenty-first century.[16],[17].

The impact of global warming on the world's weather condition to date is fairly small evaluated with what can be expected in the future, even if the raise in greenhouse gas emissions is stabilized. moreover, the results are likely to be highly irregular in their distribution, with low-lying coastal areas (such as little island developing States, the Bangladesh delta

and the Netherlands) at risk because of rising sea levels; sub-Saharan Africa at risk because of desertification; a growing number of environmental refugees; and improved pressure on sources of fresh water and on vulnerable ecosystems such as coral reefs, tundra and coastal wetlands.[16],[18].

2.7. Link Between ICT and Weather Forecasting

Weather change endangers the quality and availability of human resources. Adequate weather forecasting information can be passed through radio Broadcast or TV networks supplemented by text messages to offer alerts to the communities at risk, but access to the internet is needed to provide reference information, datasets and adaptation plans. wireless phone networks with internet browsing capability are being extended to bring adaptation forecasting information to remote communities.

Monitoring environmental and soil conditions using ICTs makes farming more profitable and sustainable. Better water management[18] using ICTs improves the overall effectiveness of water use, providing significant savings and a more sustainable use of water resources [14] The foremost areas where ICT could play a pivotal role in water management and weather forecasting are show below: See figure 1.

3. METHODOLOGY

3.1 Using Telecommunication Network For Weather Forecasting.

Extensive weather station networks are desired for monitoring key weather parameters such as wind speed, precipitation, barometric pressure, soil moisture, wind direction, air temperature and relative humidity. These parameters may be used together for forecasting and for weather modeling. The technologies needed comprise weather satellites and both local and remote automated weather stations. Just as with telecommunications networks in general, there are logistical and financial problems in achieving satisfactory global coverage to collect the required data.[14]

For Satellite observation, visible spectrum cameras are used to detect storms and deforestation. Infrared cameras are for discovering cloud and surface temperatures, sea level rise and particle detectors of solar emissions. The Geostationary Operational Environmental Satellites [4][12], and others, are able of making these observations, which are essential in providing input to weather forecasting and weather change models. Emphasis is on improving coverage of space and land based sensors. Fine resolutions are needed, with regular updates, to provide the most accurate forecasts. For instance the European Meteosat-8 located over the Atlantic Ocean at 0o longitude provides an operational European 'rapid scan' mode service, which initiated in the second quarter of 2008 (with images of Europe every 5 minutes)[13]. Meteosat-9 also at 0o offers the main full earth imagery service over Europe and Africa (with images every 15-minutes). More work is needed to set up whether Africa and other developing regions could benefit from dedicated weather satellites, with enhanced resolution over their regions, to match the standards of weather change forecasting in developed regions[12].

Weather forecasting information are Broadcasted through wireless phone networks ,bringing solution to the problems of providing sufficient mobile base station coverage and land based weather stations[15] Up to 5,000 automatic weather stations are needed to be installed at cellular sites across Africa. This will increase broadcasting of weather information via wireless phones to users and communities, including remote farmers and fishermen.

Supercomputers are being used for weather Forecasting, for accuracy in weather (general circulation) modeling, is being continuously improved through better understanding of the

basic science (including the impacts of clouds). For instance, advances in the technology, as observed by Moore's law, whereby the processing power of computers adds every two years, and more data gathering through weather and environmental sensors connected to telecommunication networks. In order to make appropriate information available to local communities these models are being extended to predict changes in regional and local weather and sea level extremes.[14].

4. DISCUSSION

This work discussed the major causes of weather change, signs of weather change and its effects. The function of information and communication technologies in mitigating, adapting to and monitoring weather change was also discussed. From the above review the major contributor to weather change is human activity notably the emission of green house gases into the atmosphere. Therefore ICTs have huge potential to mitigate, adapt to and monitor weather change in developing countries despite the fact that these countries are facing a lot of challenges that are making it difficult to apply some of the proposed solutions.

5. CONCLUSION

This paper examined the linkage between ICTs and Weather Forecasting, the signs and effects of weather condition and the ICT solutions. The role of information and communication technologies in mitigating, adapting to and monitoring weather change was also discussed.

From the above discussion it is clear that the major contributor to weather change is human activity: the emission of green house gases into the atmosphere.[2].

The paper reached the conclusion that ICTs is a corner stone which have huge potential to mitigate, adapt to and monitor weather change in developing countries even though these countries are facing a lot of challenges that are making it difficult to implement some of the proposed solutions.

In the area of weather change adaptation, ICT based strategies include,

i) Infrastructure innovation which is aimed at reducing energy consumption and Green House gases (GHGs),

ii) behavioral change and green enablement which focuses on need for global measurement and tracking of carbon reduction, as well as tools that impact positive behavioral change including software tools for measuring carbon footprint, and the use of innovative technologies and opportunities that reduce travel and transportation and

iii) Energy efficiency of data centers, electronic devices and solutions. As regards to weather change monitoring, ICTs can be used for weather forecasting, weather monitoring, predicting, detecting and mitigating the effects of natural disasters. Some of the ICTs that can be used in this area include

iv) The Global Observing System (GOS) which provides observations of the atmosphere and the Earth's surface (including oceans) from the globe and from outer space,

v) The Global Telecommunication System (GTS) which are radio and telecommunication networks for real-time exchange of a huge volume of data between meteorological centers and

vi) The Global Data Processing System (GDPS) where many networked mini, micro and supercomputers, processes an massive volume of meteorological data and generates warnings and forecasts.

6 RECOMMENDATIONS

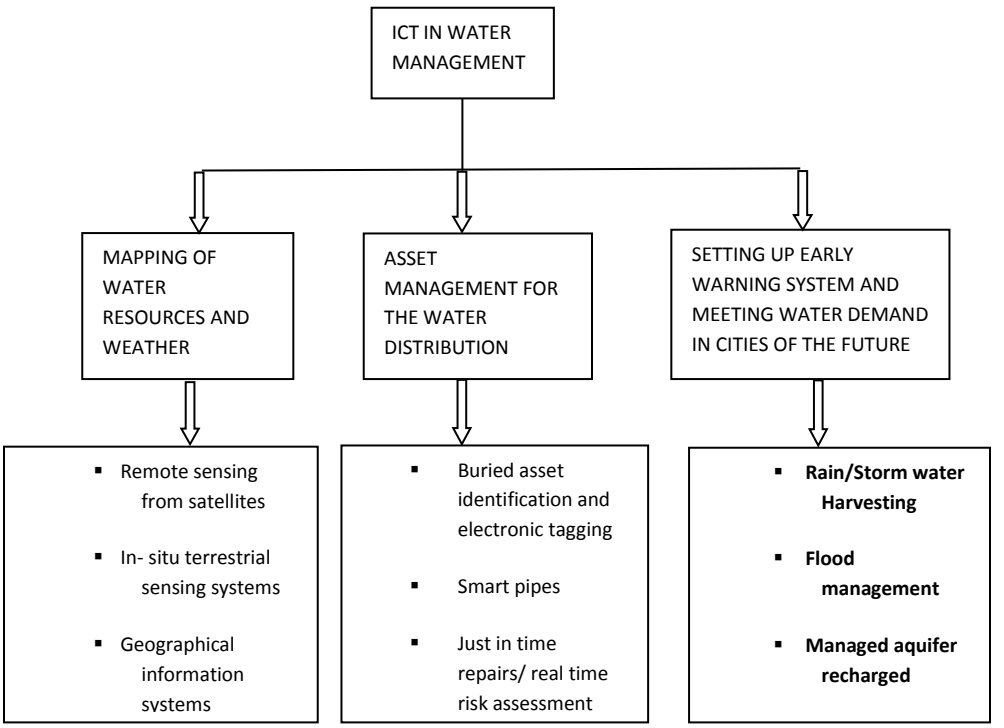
Weather change is global problem which implies that global solution which is based on use of ICT is required via:

- Addressing Weather problems at national level by each and every country through creating a budget for weather change. Developing countries that do not have the needed ICT infrastructure can improve through the use of the money each country budget's for weather change. On setting up the infrastructure, it is recommended that each government must use green technologies only.
- Mitigating weather change, all involved in the ICT sector, the education sector and the media must take an active role in disseminating information not only about the role of ICTs in mitigating weather change but also about the need for every person to reduce emissions of GHGs into the atmosphere.
- Educational administrators and institutions must include weather change in their curriculum from the lowest level possible to the highest level.
- Further, the ICT sector should develop affordable software tools that can measure the carbon foot print since it is expensive to import software which was developed elsewhere.
- All private and public owned companies should have software tools which measure its carbon foot print. This is very important because it is difficult to reduce the CO2 emissions if we cannot measure the carbon footprint,

7. REFERENCE

[1] Hartványi, T., Nagy 2009." In-sourcing Model for Food Storage and Forwarding." Acta Technica Jaurinensis, Series Logistica. Vol. 2. No. 3.

[2] Zanamwe and Okunoye 2013." Role of information and communication technologies (ICTs) in mitigating, adapting to and monitoring climate



'Figure 1': ITU-T Technology Report, Major Areas for ICT in Water Management [6].

change in developing countries " International Conference on ICT for Africa 2013, February 20 -23.

[3]http://www.wmo.int/pages/mediacentre/press_releases/pr_864_en.html

[4]http://www.itu.int/themes/climate/events/cop16/sg.html

[5] http://www.itu.int/publ/R-HDB-45/en

[6] ITU-T Technology Report, Major Areas for ICT in Water Management available at: http://www.itu.int/oth/T2301000010/en.

[7] Matarira, C.H., Makadho, J.C., Mukahanana-Sangarwe, M 2004. "Vulnerability And Adaptation Of Maize Production To Climate Change In ZIMBABWE", Zimbabwe Government Publication.2004.

[8]The Hyogo Framework for Action 2015." Building the Resilience of Nations and Communities to Disasters, available at: http://www.unisdr.org/eng/hfa/hfa.htm, 2005-2015

[9] Russell J, Zimbabwe. (2008)." Country At Risk of Climate Change Effects", available at: http://www.zimbio.com/Zimbabwe/articles/263/Zimbabwe+Country+Risk+Climate+ Change Effects.

[10] Mobile cellular subscriptions per 100 inhabitants, 2010 available at: http://www.itu.int/ITUD/ict/statistics/material/graphs/2010/Mobile_cellular_00-10.jpg.

[11] IPCC Climate Change, Synthesis Report 2007.. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change 2007.

[12] ITU/WMO Handbook "Use of Radio Spectrum for Meteorology: Weather satellite, Water and Climate Monitoring and Prediction", available at:http://en.wikipedia.org/wiki/Weather_satellite.

[13]The Nairobi Framework: http://cdm.unfccc.int/Nairobi_Framework/index.html

[14]'Wireless Sensor Networks for marginal farming in India' by Jacques Panchard, École Polytechnique Fédérale de Lausanne, Switzerland. http://commonsense.epfl.ch/Resources/thesis.pdf

[15]WMO 'Mobile Systems to Revolutionise African Weather Monitoring www.wmo.int/pages/mediacentre/press_releases/pr_855_en.html

[16] GSMA 'Green Power For Mobile' http://www.gsmworld.com/our- work/mobile planet/green power for mobile/index.htm

[17]http://ourworld.unu.edu/en/how-things-work-environmental-satellites/

[18] ITU-T Technology Report 2010. On "ICT as an Enabler for Smart Water Management (October 2010)". http://www.itu.int/oth/T2301000010.

International Journal of Computer Applications Technology and Research
Volume 6–Issue 3, 130-132, 2017, ISSN:-2319–8656

An Extension of RETRO Framework: Translating SQL Insert, Update and Delete Queries to SPARQL UPDATE

Abatal Ahmed
University Hassan,
Faculty of Science
and Technologies,
Settat, Morocco

Alaoui Khadija
University Hassan,
Faculty of Science
and Technologies,
Settat, Morocco

Larbi Alaoui
International
University of Rabat,
Sala Al Jadida,
Morocco

Bahaj Mohamed
University Hassan,
Faculty of Science
and Technologies,
Settat, Morocco

***Abstract*—**in this paper we propose a set of algorithms to improve the functioning of RETRO framework, these algorithms will convert SQL queries like INSERT, UPDATE and DELETE to the equivalent queries SPARQL Update.

Keywords*—RDB, RDF, SQL, SPARQL, Query Translation*

1. INTRODUCTION

RDF (Resource Description Framework) which was standardized by the W3C is a language for describing the semantics of data that allows sharing of its meaning between different applications. RDF provides a powerful data model based on representing data in RDF graphs that can be queried using SPARQL. SPARQL (SPARQL Protocol and RDF Query Language) was proposed and standardized by W3C as a query language for RDF.

2. RETRO FRAMEWORK

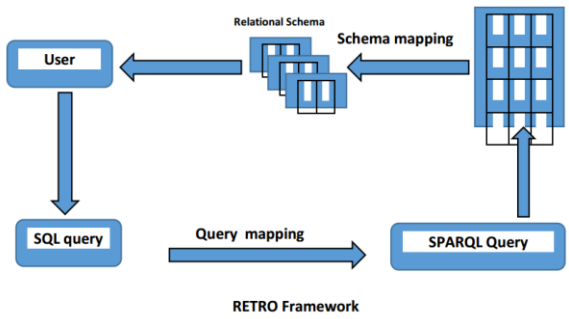


Figure 1. RETRO Schema and Query Mapping

RETRO Framework is to conserve and translate SQL queries into SPARQL. This framework is composed of several algorithms.

The first *schema-mapping ()* algorithm is to convert an RDF store RDB, this algorithm uses the property table method (for every single predicate it creates a table with two columns (subject and object). This algorithm returns a map P representing the relational schema for users, this relational schema is generated by extracting each predicate name from the map P, S and O are added to each predicate.

Example : **RDF Store**

(S1,name,ali) (S3,name,ahmed) (S3,age,27)

(S1,age,25) (S2,age,30) (S2,name,mohamed)

(S1,phone,555 0123) (S1,phone,666 2156) (S1,name,ali)

(S3,phone,999 3453) (S1,name,ali) (S1,name,ali)

RDBMS :

Table : **name**

S	O
S1	ali
S2	Mohamed
S3	ahmed

Figure. 2 Example of an RDB Table

Before starting to translate queries, RETRO separates SQL clauses, so that an algorithm as specified must convert each clause.

The *TransSqlFromClause ()* algorithm that will return triple patterns in the SQL FROM clause. Then *TransSqlWhereClause ()* algorithm that will translate into SPARQL WHERE clause WHERE clause, using pattern returned by *TransSqlFromClause ()*. After the algorithm Trans SQL Select clause () also translate the SQL select clause in select SPARQL.

Then a main algorithm *query-mapping ()* that brings together the outputs of these sub-programs, to combine and give a query SPARQL Equivalent to the SQL.

3. RETRO UPDATE QUERIES

In this section we give an extension to the Framework retro, our proposal is to convert SQL queries type INSERT / DELETE / UPDATE to equivalent SPARQL UPDATE queries.

Our algorithms are based on the *query-mapping ()* algorithm [1] *query-mapping ()* is to convert an RDF STORE RDB, this algorithm property table method, for every single predicate it creates a table with 2 columns (subject and object).

SQL Query	SPARQL Query

delete from name where subject='S1'	DELETE DATA {
	's1' name o
	}

delete from name where object='Mohammed'	DELETE DATA {
	s name 'mohammed'
	}

3.1 Insert Query

To insert the table name in the query is executed

Insert into name (subject, object) values ('S1', 'Ali').

Our algorithm *transSqlInsert ()*, receives the SQL query string form, and it will extract the triple patterns to generate an equivalent SPARQL INSERT

Therefore, the algorithm takes the table 'name' as a predicate for the new triple pattern and the values corresponding to the two columns (subject, object) subject and object to have a triple subject predicate object.

INSERT DATA {'S1' name 'Ali' }

- The algorithm uses the method *splitQuery ()* to cut the SQL query and extract the name of the table matches the predicate, the values of subject and object.
- Then do call the *generate ()* method that will deduct the triple pattern from the result *splitQuery ()* and creating an equivalent SPARQL INSERT query.

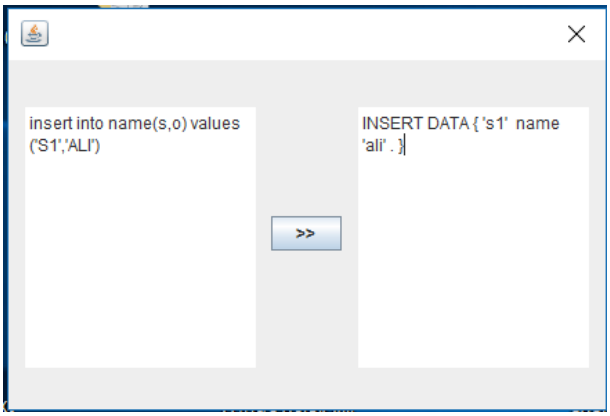


Figure. 3 Example of an SQL Insert converted

3.2 Delete Query

To delete from the table name it executes the query

Delete from name Where subject = 'S1';

TransSqlDelete () receives the SQL query string form, and it will extract the triple pattern to generate an equivalent SPARQL DELETE.

ALGORITHM: ***TRANSSQLDELETE ()***

Input	Q an SQL Insert Query
Output	Z a SPARQL Delete Query

```
1  p<= ''
2  tps<=null
3  o<= ''
4  q=splitQuery(Q) ;
5  p=q.getPredicatValue() ;
6  tps=q.getTriples() ;
7  Z =generateQuery(p,tps) ;
8  Return Z ;
```

This algorithm begins by extracting the name of the table representing the predicate of GDI drawing propose, and extract the Boolean condition in the WHERE clause to build a triple pattern is using *spliteQuery ()* method, which receives as a parameter the SQL DELETE as a

string. After extracting the predicate for the request, the method *generateQuery (p, tps)* reference SPARQL query Delete, from the extracted variable (p predicate and list of pattern triples).

Figure. 4 Example of an SQL Delete converted

3.3 Update Query

In the SPARQL language there is no equivalence of the update query, so to execute an update query, you must combine the two DELETE queries then INSERT, but you must keep the delete values to re-insert them in the RDF without loss of data

4. CONCLUSION

In this paper, we described an extension of RETRO framework, this extension is in the form of n set of algorithms to convert SQL queries INSERT, DELETE and UPDATE to SPARQL Update, and these algorithms are implemented and tested by java language. Another promising avenue for future work is to implement a framework for translating queries for other storing methods.

4. REFERENCES

[1] Jyothsna Rachapalli, Vaibhav Khadilkar, Murat Kantarcioglu and Bhavani Thuraisingham, "RETRO: A Framework for Semantics Preserving SQL-to-SPARQL Translation", EvoDyn Workshop, 2011.

[2] Bajda-Pawlikowski, K.: Querying RDF data stored in DBMS: SPARQL to SQL Conversion. Technical Report TR-1409, Yale Computer Science Department, USA

[3] A. Seaborne, G. Manjunath: SPARQL/Update: A language for updating RDF graphs. External HPL-2007-102, HP Labs Bristol (2007).

[4] W3C, Resource description framework (RDF): concepts and abstract syntax, in: G. Klyne, J.J. Carroll, B. McBride (Eds.), and W3C Recommendation [S], 10 February2004. http://www.w3.org/TR/2004/REC-rdf-concepts- 20040210

International Journal of Computer Applications Technology and Research
Volume 6–Issue 3, 133-140, 2017, ISSN:-2319–8656

Classification Model to Detect Malicious URL via Behaviour Analysis

N.Jayakanthan
Department of Computer Applications
Kumaraguru College of Technology
Coimbatore, India

A.V.Ramani
Department of Computer Science
Sri Ramakrishna Mission Vidyalaya College of Arts and Science
Coimbatore, India

Abstract: The challenging task in cyber space is to detect malicious URLs. The websites pointed by the malicious URLs injects malicious code into the client machine or steals the crucial information. As detecting a phishing URL is a challenging task, it is essential to enhance detection techniques against the emerging attacks. The most of the existing approaches are feature based and cannot detect dynamic attacks. Mostly the attacker uses the input form, active content and embeds @ symbol in URL for malicious attack. To detect this attack, a Behaviour based Malicious URL Finder (BMUF) algorithm is proposed. It analyzes the behaviour of the URL. The FSM based state transition diagram is used to model the URL behaviour into various states. The state transition from initial to final state is used for classification. This approach tests the genuine and malicious behavior of the URL based on the responses to the user. It accurately detects the nature of the URL.

Keywords: Malicious URL, Behaviour based Malicious URL Finder, Finite State Machine, Input form, Active Content

1. INTRODUCTION

The Malicious URL leads the user to phishing websites. These websites steal the user's confidential information without their knowledge using fake information form, active contents, and embed @ symbols in URL. These attacks inject malicious code in the client machine, and it controls the machine and spreads the malicious code to other machines in the same network[22]. The malicious web sites resemble the websites of the trusted organizations such as banks, government agencies, and e-commerce websites.

Generally most of the phishing attacks are Drive-by-downloaded attack. It installs the malicious code in users system to generate attack [7]. The code is automatically downloaded from the web page of the attacker without the permission of the user. This behavior is an important feature to detect web attacks.

The URL redirection mechanism is commonly used to carry out web attacks. The attacker redirects the visitor to the malicious website [15].The attacker r performs the following activities to make a successful attack. They are developing fraudulent websites and motivating users to visit those sites through malicious URL. The @ symbol is used to embed a malicious URL with a genuine URL. Apart from that input form, active contents also redirect the user to the malicious websites.

A number of approaches have been developed in recent years to detect the malicious attacks.

These include detecting suspicious websites [10], educating and training users [12], white list and black list based fault detection and feature based analysis of legitimate and malicious URLs.

Most of the web browsers are having built-in phishing detection abilities based on white and black lists. There exists no testing approach for anti-phishing professionals to manually verify suspected URL and intimate the administrators to take down the fake URLs. More over the phishers can exploit the cross site scripting (XSS) vulnerabilities by generating forms, active contents and @ symbol, motivating us to device behaviour based testing approach for malicious URL detection.

The proposed approach detects the malicious URLs based on the behaviour. Most of the existing approach detects the malicious URLs using lexical and host based features. But attacks in present scenario are highly dynamic which is not detectable through feature analysis. So we propose a behaviour based approach to detect the malicious URL.

The contribution of the proposed approach is as follows.

- It is a dynamic approach that detects the malicious URLs based on their behaviour.
- Behaviour based Malicious URL Finder algorithm is developed to detect the nature of the URL.
- FSM based state transition diagram is developed to model the URL behaviour in various states
- It improves the accuracy of the classification
- It is a light weight approaches capable of detecting malicious URL with low performance overhead.

The paper is organized as follows. Section 2 describes the related work done for malicious URL detection. Architecture of the proposed system is given in section 3. Section 4 deals with methodology. Section 5 discusses the analysis of the URL. Finally section 6 concludes the paper.

2. REVIEW OF RELATED WORK

Hossian Shahriar and Mohammed zulkernine [10] developed a tool phishTester to test the trustworthiness of the website

based on the behavior of the web application. They used Finite State Machine that captures the submission of forms with random inputs and corresponding responses.

Hyunsang Choi et al[11] analyze various types of discriminative features acquired from lexical, webpage, DNS, network, and link popularity properties of the associated URLs. The used SVM to detect malicious URLs, and both RAkEL and ML-kNN were used to identify attack types.

Sidharth Chhabra[6] et al and Y. Alshboul et al[1] found some malicious attacks obfuscating the host with largest host names, another domains and misspelled various. All these attacks hide the malicious URL behind the genuine URL. It leads the user to the malicious website.

Cheng Cao and James Caverlee[4] proposed a method to identify the malicious URLs through posting based features and click based features. The behavioral signals are analyzed for classification and this method yields 86% accuracy. Few machine learning approaches extract the URL features to train the classification model through training data. The features are categorized in to two classes- static and dynamic. In static analysis [4][13][14][2], the information is analyzed without the execution of the URL, but in dynamic approach the run time behavior of the URL is used for classification.

Charmi Patel and Hiteishi Diwanji[5] analyze the lexical and network based features using URL pattern matching algorithm. This algorithm analyzes the different patterns of URL to detect the malicious one. R.K. Nepali et al[16] use four machine learning algorithms - Naïve Bayes, random forest, support vector machine, and logistic regression to detect malicious URL.and obtain an accuracy of 97% using random forest algorithm. Y. Tao [21] proposed a dynamic method which mines the internet access log file to detect the malicious activity.

Peilin Zhao and Steven C H Hoi [18] proposed a Cost-Sensitive Online Active Learning (CSOAL) frame work to detect malicious URLs. The experimental results proved the efficiency of algorithm in classification. The black list based approaches [20][3][9] detects the URLs using the blacklisted profile. But they are incapable to detect emerging attacks.

H.K. Pao[17] et al method calculates Conditional Kolmogorov Complexity of the URL's with reference to genuine and malicious URLs. It compares the given URL with malicious or genuine URL databases for classification. W. Chu et[8] proposed a phishing detection method based on machine learning approach. The lexical and domain based features are analyzed for classification. This method properly classifies even the changes in the phishing URLs. E. Sorio[19] proposed a method to detect the hidden URLs based on their lexical features. Nearly 100 URLs are analyzed and experimental results show the efficiency of this approach

3.ARCHITECTURE OF THE PROPOSED SYSTEM

The architecture of the proposed system is given in figure 1. The components are browser, Behavioural Extraction, FSM Model, BMUL Classifier and Final classification.

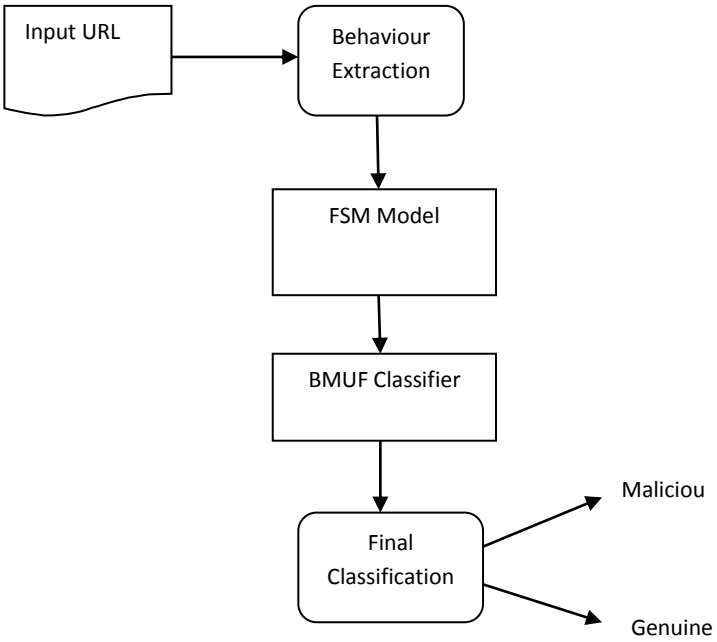


Figure 1.Architecture Diagram

available, try the font named Computer Modern Roman. On a Macintosh, use the font named Times. Right margins should be justified, not ragged.

3.1Browser:

The URL is the input to the browser. The behaviour of the URL is extracted for the analysis.

3.3 FSM Model:

FSM state transition diagram model the URL behaviour in two various states. The states are derived from 3 inputs and 13 responses. The transition from the initial to final state leads to the classification.

3.3. BMUF Classifier:

The Malicious URL Finder (MUF) is rule based algorithm which analyzes the URL through using FSM state transition diagram. If any malicious behaviour is detected it marks the URL as malicious and collects the behaviour and reports it to the user.

4. METHODOLOGY

The proposed method uses the Behaviour based Malicious URL Finder (BMUF) algorithm to analyze the behavior of URL to detect whether the URL is genuine or malicious. The FSM based state transition diagram is used capture the URL's behavior in various states. The state transition from initial to finial state classifies the natural of the URL. This classifier improves the accuracy of the malicious URL Detection.

4.1 Algorithm:

Behaviour based Algorithm Malicious URL Finder (BMUF)

//Input: URL of the Webpage

//Output: Genuine or Malicious

M_B=Null // set of malicious behaviour.

Step1 : Consider the Input URL

If (Automatic content (ad) download occurs) then

Set Status = Malicious

$M_B = M_B$ U ad

Step 2 : Check the webpage pointed by the URL contains input form

When the user submits input form

a. If the user is redirected to a new webpage and URL of the page contain malicious words (mw) then

Set Status = Malicious

$M_B = M_B$ U mw

b. If the user is redirected to a new webpage and content is automatically downloaded (ad) then

Set Status = Malicious

$M_B = M_B$ U ad

Step 4 : Check the webpage pointed by the URL contains active content/s(ac) When the user access the active content then

a. If the user is redirected to a new webpage andURL of the page contain malicious words(mw) then

Set Status = Malicious

$M_B = M_B$ U mw

b. If the user is redirected to a new webpage and content is automatically downloaded (ad) then

Set Status = Malicious

$M_B = M_B$ U ad

Step 5: Check the webpage pointed by the URL contains @ symbol

a. It redirects the user to a webpage and URL of the webpage contain malicious word/s(mw) then

Set Status = Malicious

$M_B = M_B$ U mw

b. If the user is redirected to a new webpage and content is automatically downloaded (ad) then

Set Status = Malicious

$M_B = M_B$ U ad

Step 6 :

If status="Malicious" Then

Display URL is malicious

Display Set of Malicious behaviour M_S

Else

Display URL is genuine

4.2 FSM Model

The behaviours of the URL are modeled using Finite State Machine (FSM). Various symptoms of malicious and genuine URLs for FSM are developed based on submission of the information window, accessing active content. The norms are established by our literature survey. The malicious behaviours are identified as follows

a. Malicious content automatically downloaded from Web page of the URL

b. User access the input form or active content which leads to another webpage where content is automatically downloaded from the webpage.

c. Malicious Word or @ symbol in the URL.

The FSM is represented by $<Q, \sum, q_0, \delta, F>$ where F is the finite set of states, q_0 is the initial state, $\sum$ is the finite set of inputs, δ is the state transition function, and F is the set of final states.

(i) Q is a finite nonempty set of states. q_0 to q_{13} that represents the various behavioral states of URL.

(ii) $\sum$ is finite non empty set of inputs called an input alphabet. It is a combination of test cases $<I , K_i>$

(iii) δ is a function which maps $Q * \sum$ into Q and is usually called direct transition function. This is the function which describes the change of the state during transition. The mapping is usually represented by a transition table or transition diagram. The transition represents the behavioral change of the URL.

(iv) q_0 is the initial state. It represents the initial stage of the URL.

(v) F is the set of final states. It is assumed here that there may be more than one final state. The final states characterize the genuine or malicious behavior of the URL.

$\sum = <I_0,K_1>\ldots.<I_1,K_n>$ is the set of input symbols. Let q_0 be the initial state of the machine. It represents the input URL. The state q_1 of the machine for the input $<I_0,K_1>$ is as follows.

$q_1 = \delta(q_0 , < I_0,K_1>) = \delta 1(q_0 , < I_0,K_1>)$ where $\delta = \delta 1 : Q \times \sum$

The change in the state due to the second input symbol $< I_0,K_2>$) is q2.

$q_2 = \delta(q_1,< I_0,K_2>) = \delta(\delta 1(q_0, < I_0,K_1>),<I_0,K_2>)$

$= \delta_2 (q_0 , < I_0,K_1><I_0,K_2>)$ (1)

Where $\delta_2 : Q * I^2 \rightarrow F$

The function of the FSM is defined as follows

$q_n = \delta_n (q_0 , <I_0,K_1>.....<I_n,K_n>)$

$= \delta(\delta_{n-1} (q_0, , < I_0,K_1 >.....<I_{n-2} , k_{n-2} >,< I_{n-1} , k_{n-1} >)$ (2)

Equations 1 and 2 show the mapping function from one state to another state in the proposed approach

The FSM model is represented as a set of inputs (I_0 to I_2) and corresponding responses (K_0 to K_{13}) are discussed in detail in the following paragraphs. A URL is classified as malicious or benign based on the traversal from initial state to final state. The state transition is given in figure 2. The final stages (q_2, q_5, q_7, q_8, q_{11}, q_{13}) are legitimate and some of the final states (q_4, q_6, q_9, q_{10}, q_{12}) are phishing.

A state transition occurs for a given input and subsequent response. The transition is represented as <Input, response> pair as shown in the figure 2. The pair $<I_2, K_2>$ represents the input I_2 and its corresponding response K_2. There are three kinds of inputs.

1. The input URL (U),
2. URL leads to a webpage which contains malicious Active content.
3. URL leads to a webpage which contains input form

[Example:https://www.perspectiverisk.com/wp-content/uploads/2016/09/Login.png]

The features that represent the set of responses are given below.

iw : Indicates the user fills the information window and submits it.
ac : The user access the active content
@ : The presence of @ symbol in the URL
re : The page is redirected. It may happen due to the submission of input form or response of user interaction with active content or malicious domain pointed by the @ symbol.
mw : URL contains malicious word
p : Presence of the redirected page.
d : The content gets automatically downloaded from URLs web page. They are counterfeit executable programs.

These features are used to classify whether the URL is phishing or genuine. The ! symbol represents the absence of a particular feature (!iw represents the absence of the information window).

The proposed approach distinguishes the malicious URL from the legitimate one based on the behavior of the URL. The input and responses are given in the table1.

Table 1. Input and responses

Name	Representation	Explanation
I_0	U	Input URL
I_1	AC	The URL leads to a webpage that contains active content
I_2	I	The URL leads to a webpage that contains Input form.
K_1	!iw !ac !@ !re !mw !p !d	No Information window, no active content, no @ symbol present in URL, no redirection, no malicious word, no redirected page , no automatic content download.
K_2	iw !ac !@ !re !mw !p !d	User submit information window, no active content, no @ symbol present in URL, no redirection, no malicious word, no redirected page , no automatic content download.
K_3	iw !ac !@ re mw p !d	User submit information window , no active content, no @ symbol present in URL, redirection occurs, malicious word present in the URL, redirected page present, no automatic content download.
K_4	iw !ac !@ re !mw p !d	User submit information window , no active content, no @ symbol present in URL, redirection occurs, no malicious word, redirected page present, no automatic content

		download .
K_5	iw !ac !@ re !mw p d	User submit information window, no active content, no @ symbol present in URL, redirection occurs, no malicious word , redirected page present, automatic content download occurs.
K_6	!iw ac !@ !re !mw !p !d	No Information window present, active content occurs, no @ symbol present in URL, no redirection occurs, no malicious word, no redirected page present, no automatic content download
K_7	!iw ac !@ re !mw p !d	No Information window present, active content occurs, no @ symbol present in URL, redirection occurs, no malicious word , redirected page present, no automatic content download
K_8	!iw ac !@ re mw p !d	No Information window present, active content occurs, no @ symbol present in URL, redirection occurs, malicious word present in the URL, redirected page present, no content download occurs.
K_9	!iw ac !@ re !mw p d	No Information window present, no active content, no @ symbol present in URL, redirection occurs, no malicious word, redirected page present, automatic content download occurs. .

K_{10}	!iw !ac @ !re !mw !p !d	No Information window present, no active content, @ symbol present in URL, no redirection occurs, no malicious word, no redirected page present, no content download occurs. .
K_{11}	!iw !ac @ re mw p !d	No information window present, no active content, @ symbol present in URL, redirection occurs, malicious word present in the URL , redirected page present, no downloads occurs.
K_{12}	!iw !ac @ re !mw p !d	No information window present, no active content, @ symbol present in URL, redirection occurs, no malicious word, redirected page present, no downloads occurs.
K_{13}	!iw !ac @ re !mw p d	No information window present, no active content, @ symbol present in URL, redirection occurs, no malicious word, redirected page present, automatic content download occurs.
K_{14}	!iw ! ac !@ !re !mw !p d	No information window present, no active content, no @ symbol present in URL, no redirection occurs, no malicious word, no redirected page present, automatic content download occurs.

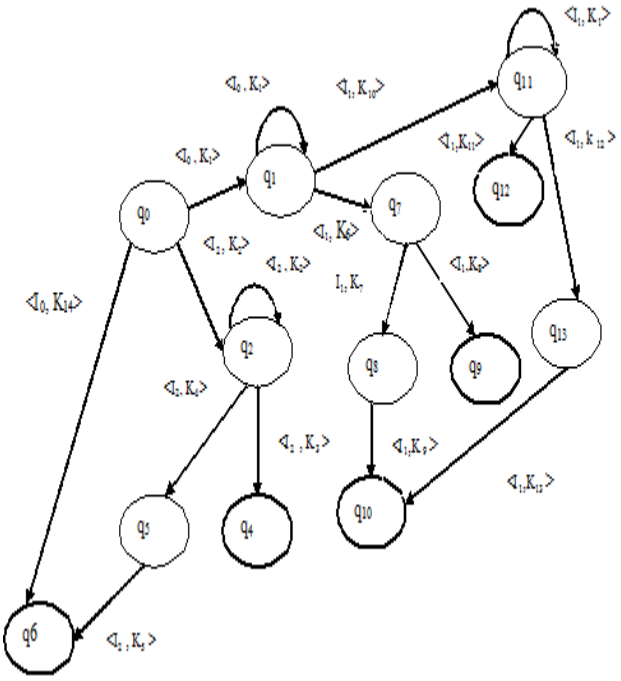


Figure 2 FSM state transition diagram

Figure 2 shows the state diagram of the FSM model. Here q_0 is the initial state. If the page downloaded from the given URL contains no information window, no active content, no @ symbol, no malicious word , no redirection and no automatic content downloads (code injection), then the next state is considered q_1. The finite set of states is as follows($q_1,q_2,q_3,q_4,,q_5,q_6,q_7,q_8,q_9,q_{10},q_{11},q_{12},q_{13}$). Each state represents the behavioral state of the URL.

4.3. State transition

The traversals of the URL help to identify it as malicious or legitimate. If the URL is having the state sequences given in table2, then it is considered as malicious (malicious final states are shown in a dark circle in the figure2.

Table 2 List of malicious states

States	Test cases	Description
q0 , q6	$<I_0,K_{14}>$	The content is automatically downloaded from the web page of the URL.
q0, q2, q4	($<I_2,K_2>$,$<I_2,K_3>$)	The user fills the information window and clicks the submit button, The user is redirected to a new page where the URL contains malicious word.
q0,q2,q5,q6	($<I_2$, $K_2>$, $<I_2$, $K_4>$, $<I_2$, K_5 >)	The user fills the information window and clicks the submit button, the user is redirected to a new page, and the contents are automatically downloaded from that page.
q0,q1,q7, q8,q10	($<I_0$, $K_1>$, $<I_1$, $K_6>$, < I_1, K_7 > , $<I_1,K_9>$)	The user accesses the active content, it leads the user to a new web page and the contents are automatically downloaded from the page.
q0,q1,q7,q9	($<I_0$, $K_1>$, $<I_1$, $K_6>$, $<I_1,K_8>$)	The user accesses the active content, it leads the user to a new web page where the URL contains malicious word
q0,q1, q11, q12	($<I_0$, $K_1>$, $<I_1$, $K_{10}>$, $<I_1,K_{11}>$)	The URL contain @ symbol it redirect the user to a another webpage it contains malicious word in the URL
q0,q1,q11,q13, q10	($<I_0$, $K_1>$, $<I_1$, $K_{10}>$, $<I_1$, K_{12} >, $<I_1,K_{13}>$)	The URL contains @ symbol that leads the user to a new web page and the contents are automatically downloaded from the page.

5.ANALYSIS OF THE URL

The set of URLs used for analysis are given below

1.https://www.perspectiverisk.com/wp-content/uploads/2016/09/Login.png

2.http://demo.smartscreen.msft.net/other/explotframe.html

3.www.yahoo.com

4.https://phishme.com/macro-based-anti-analysis/acf.css

5.http://geniune.com@malicious.com/config/change.html

The proposed Behaviour based Malicious URL Finder (BMUF) algorithm analyze the behaviour of each URL using various rules and classify it as genuine or malicious. The behavior and classification are given in the following table3.

Table 3 BMUF Classification

URL.NO	URL	Input Form	Active Content	@ Symbol	Redirection	Malicious Word	Auto Content download	Classification
1	Y	Y	N	N	Y	N	Y	Malicious
2	Y	N	Y	N	Y	Y	N	Malicious
3	Y	N	N	N	N	N	N	Genuine
4	Y	N	Y	N	N	N	Y	Malicious
5	Y	N	N	Y	N	Y	N	Malicious

Finite State Machine (FSM) model the behavior as various states. The state transition from the initial state to final state is used for classification. The state transition for each URL is given in table 4.

Table 4 State Transition

URLNO	States	Test case	Description	Classification
1	q0 , q6	$<I_0,K_{14}>$	The content is automatically downloaded from the web page of the URL.	Malicious
2	q0,q1,q7,q9	($<I_0$, $K_1>$, $<I_1$, $K_6>$, $<I_1,K_8>$)	The user accesses the active content, it leads the user to a new web page where the URL contains malicious word	Malicious
3	q0,q1	($<I_0$, $K_1>$)	URL is present but no malicious activities detected.	Genuine
4	q0,q1,q7, q8,q10	($<I_0$, $K_1>$, $<I_1$, $K_6>$, $<I_1$, $K_7>$, $<I_1,K_9>$)	The user accesses the active content, it leads the user to a new web page and the contents are automatically downloaded from the page.	Malicious
5	q0,q1, q11, q12	($<I_0$, $K_1>$, $<I_1$, $K_{10}>$, $<I_1,K_{11}>$)	The URL contain @ symbol it redirect the user to a another webpage it contains malicious word in the URL	Malicious

The classification of the list of URL is given in table 5.

Table 5 Classification

URL. No	URL	Classification
1	https://www.perspectiverisk.com/wp-content/uploads/2016/09/Login.png	Malicious
2	http://demo.smartscreen.msft.net/other/explotframe.html	Malicious
3	www.yahoo.com	Genuine
4	https://phishme.com/macro-based-anti-analysis/acf.css	Malicious
5	.http://geniune.com@malicious.com/config/change.html	Malicious

6. CONCLUSION:

The web attacks are challenging problem for the web users. Detecting malicious URL is a complex task due to the dynamic behavior of the URL. The proposed classification model to detect malicious URL is based on the behavior. The Behaviour based Malicious URL Finder (BMUF) algorithm analyzes the behavior in sequence of steps to detect the URL is genuine or malicious. Finite State Machine state transition diagram capture the behaviour into various states. The state transition from initial to final states leads to classification. Thirteen states are derived from 3 inputs and 13 responses. The final states represents whether a URL is genuine or malicious. The proposed algorithm improves the accuracy of the classification.

7. REFERENCES

1. Y. Alshboul, R. Nepali, and Y. Wang, “Detecting malicious short urls on twitter,” 2015.

2. Birhanu Eshete, A. Villafiorita, and K. Weldemariam, “Binspect: Holistic analysis and detection of malicious web pages,” in Security and Privacy in Communication Networks. Springer, 2013, pp. 149–166.

3. S. Bo, M. Akiyama, Y. Takeshi, and M. Hatada, "Automating url blacklist generation with similarity search approach," IEICE TRANSACTIONS on Information and Systems, vol. 99, no. 4, pp. 873–882, 2016.

4. C. Cao, J. Caverlee, Detecting spam urls in social media via behavioral analysis, in: Advances in Information Retrieval, Springer, 2015, pp. 703– 714.

5. Charmi Patel , Hiteishi Diwanji Research on Web Content Extraction and Noise Reduction through Text Density Using Malicious URL Pattern Detection",International Journal of Scientific Research in Science Engineering and Technology",Volume 2, Issue 3, ISSN : 2394-4099,May-June 2016.

6. S. Chhabra, A. Aggarwal, F. Benevenuto, and P. Kumaraguru, "Phi. sh/$ ocial: the phishing landscape through short urls," in Proceedings of the 8th Annual Collaboration, Electronic messaging, Anti-Abuse and Spam Conference. ACM, 2011, pp. 92–101.

7. S.Chitra K. S. JayanthanS. PreethaR. N. Uma Shankar "Predicate based Algorithm for Malicious Web Page Detection using Genetic Fuzzy Systems and Support Vector Machine"International Journal of Computer Applications. Volume 40 - Number 10.2012. DOI:10.5120/5000-7277.

8. W. Chu, B. B. Zhu, F. Xue, X. Guan, and Z. Cai, "Protect sensitive sites from phishing attacks using features extractable from inaccessible phishing urls," in Communications (ICC), 2013 IEEE International Conference on. IEEE, 2013, pp. 1990–1994.

9. M. Felegyhazi, C. Kreibich, and V. Paxson, "On the potential of proactive domain Blacklisting." LEET, vol. 10, pp. 6–6, 2010.

10. Hossain Shahriar and Mohammad Zulkernine Trustworthiness testing of phishing websites: A behavior model-based approach. Future Generation Comp. Syst. 28(8):1258-1271(2012).DOI 10.1016/j.future.2011.02.001.

11. Hyunsang Choi, Bin B. Zhu, Heejo Lee, "Detecting Malicious Web Links and Identifying Their Attack Types", InWebApps, June 2011.

12. D. Irani, S. Webb, J. Giffin, C. Pu, Evolutionary study of phishing, in: Proc. Of the 3rd Anti- Phishing Working Group eCrime Researchers Summit, Atlanta,Georgia, October 2008, pp.1–10.

13. J. Ma, L. K. Saul, S. Savage, and G. M. Voelker, "Beyond blacklists: learning to detect malicious web sites from suspicious urls," in Proceedings of the 15th ACM international conference on Knowledge discovery and data mining. ACM, 2009, pp. 1245– 1254.

14.J.Ma, L. K. Saul, S. Savage, and G. M. Voelker "Learning to detect malicious urls," ACM Transactions on Intelligent Systems and Technology (TIST), vol. 2, no. 3, p. 30, 2011.

15.Mitsuaki Akiyama, Takeshi Yagi, Takeshi Yada, Tatsuya Mori, Youki Kadobayashi,"Analyzing the ecosystem of malicious URL redirection through longitudinal observation from honeypots", Journal of Computer & Security, January 2017.

16.R. K. Nepali and Y. Wang, "You look suspicious!!: Leveraging visible attributes to classify malicious short urls on twitter," in 2016 49th Hawaii International Conference on System Sciences (HICSS). IEEE, 2016, pp. 2648–2655.

17.H.K. Pao, Y.-L. Chou, and Y.-J. Lee, "Malicious url detection based on kolmogorov complexity estimation," in Proceedings of the The 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology-Volume 01. IEEE Computer Society,2012, pp. 380–387.

18.Peilin Zhao,Steven C H Hoi"Cost-sensitive online active learning with application to malicious URL detection",Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining,Chicago, Illinois, USA — August 11 - 14, 2013. 919-927.[

19. E. Sorio, A. Bartoli, and E. Medvet, "Detection of hidden fraudulent urls within trusted sites using lexical features," in Availability, Reliability and Security (ARES), 2013 Eighth International Conference on.IEEE, 2013, pp. 242–247.

20. B. Sun, M. Akiyama, T. Yagi, M. Hatada, and T. Mori, "Autoblg: Automatic url blacklist generator using search space expansion and filters," in 2015 IEEE Symposium on Computers and Communication (ISCC). IEEE, 2015, pp. 625–631.

21.Y. Tao, "Suspicious url and device detection by log mining," Ph.D. dissertation, Applied Sciences: School of Computing Science, 2014.

22. Zhi-Yong,Ran Tao, Zhen-He cai and Hao Zhang Li A. "Web Page Malicious Code Detect Approach Based on Script Execution". International Conference on Natural Computation 009.308-312. DOI:10.1109/ICNC.2009.363.

International Journal of Computer Applications Technology and Research
Volume 6–Issue 3, 141-158, 2017, ISSN:-2319–8656

A Comprehensive Classification of MANETs Routing Protocols

Diaa Eldein Mustafa Ahmed
Faculty of Computer Science and Information Technology, Sudan University for Science and Technology, Sudan

Othman O. Khalifa
Department of Electrical and Computer Engineering, International Islamic University Malaysia

***Abstract*:** Mobile Ad hoc Networks (MANETs) is self-organized and self-configured networks that doesn't need any cellular infrastructure such as access point (AP), base station (BS) or fixed transmission links. Routing is point of research focus since the invention of commercialized mobile ad-hoc networks. In the literature the classical taxonomy of MANETs routing protocols based on the route computation process subdivided into: Proactive (Table -Driven), Reactive (On-Demand) and Hybrid (inherit both features of the proactive and reactive).However this classification hide other types of MANETs routing protocols play a significant roles nowadays such as(Multicast , Energy and Power- aware, Geographical routing and Hierarchical Routing). This paper provides a comprehensive study of MANETs basic taxonomy of routing protocols. In addition intense comparison between some examples of each routing protocols category has been made to further facilitate research in this area.

Keywords: MANET, Routing classification, AODV, OLSR, TORA, ZHLS, LANMAR, CGSR

I. INTRODUCTION

The origin of Mobile ad hoc network (MANET) was started in 1970 as packet radio network (PRNET), later on different researches were made on it in different ages, it become popular and gain a great deal of importance from both of the researchers and industry [1]. The Mobile Ad hoc Networks (MANET) is a set of wireless mobile nodes (Mobile phone, laptop, PDA, MP3 player and etc...) which can act as a transmitter, router or receiver. MANET is a peer-to-peer communication technique arises when a group of mobile nodes (MNs) can performed a communications through multi-hop routing using the multi-hop wireless link without centralized administration. MANET is homogeneous when the mobile nodes (MNs) are similar structure, platforms and equal capabilities and responsibilities to perform, and heterogeneous when otherwise. In MANETs mobile nodes that are free in moving in and out in the network, any new node can join the network at any time anywhere, likewise any node can leave the network. MANETs have several prominent features such as topological flexibility, robustness, rapid deployment, inherent mobility support and highly dynamic topology, fault-resilience, self-healing and independence of fixed infrastructure spark off many vision based applications[2, 3]. Routing protocols establishes the governing rules and define the set of parameters that indicate how the packets are exchanged between communicating nodes of MANETs[2] .Recently there are different routing protocols algorithms are proposed to overcome most of the MANETs challenges such as dynamic topology changes , limited bandwidth , link failure due to node mobility, limited power on mobile nodes, power consumption due to routing computation and etc…This paper provides comprehensive survey for the routing strategies of MANETs and discusses presently offered technological that enhances the overall performance of routing protocols. The rest of this paper is organized as follows. Section II gives an overview of the recent Challenges due to MANETs routing protocols. Section III presents a broadly and comprehensive classification of MANET routing protocols with examples for each one category, in addition provides comparative study between different routing protocols. Finally section IV concludes the paper and presents some future works.

II. CHALLENGES FACING ROUTING PROTOCOL DESIGN

MANET works under no fixed infrastructure in which every node works likes a router that stores and forwards packet to final destination. Routing is one of the most challenging tasks in MANETs. Due to its dynamic topology changes, limited bandwidth, limited battery power available in each node, frequent link failure, interference, limited resources and etc., amongst all these characteristics, there are some challenges that protocol designers and network developers are faced with. Therefore routing discovery and maintenance are critical issues in these networks. Here we will focus on most popular and important problems that facing the development of MANET routing protocols[4]:

1-Asymmetric links: Most of the wired networks rely on the symmetric links which are always fixed. But this is not case with MANETs networks as the nodes are mobile and constantly changing their position within network.

2-Routing Overhead: In MANETs networks, nodes often change their location within network due to frequent change in topology and high mobility. So, some stale routes are generated in the routing table which leads to unnecessary routing overhead.

3-Interference: This is the major problem with MANETs networks as links come and go depending on the transmission characteristics, one transmission might interfere with another one and node might overhear transmissions of other nodes and can corrupt the total transmission.

4-Dynamic Topology: Since the topology is not constant; so the mobile node might move or medium characteristics might change. In MANETs networks, routing tables must somehow reflect these changes in topology and routing algorithms have to be adapted. For example in a fixed network routing table updating takes place for every 30sec. This updating frequency might be very low for MANETs networks.

5-Distributed operation: With no central hierarchy of routers, routing must be distributed amongst the participant nodes.

6-Loop-freedom: Aim to avoid route discovery or maintenance processes from spinning from node to node indefinitely.

7-Demand-based operation versus Proactive operation: To minimize the control overhead in the network and thus not waste the network resources (bandwidth, battery, memory, etc...) the protocol should be reactive. This means that the protocol should react only when needed and that the protocol should not periodically broadcast control information.

8-Unidirectional link support: The radio environment can cause the formation of unidirectional links. Utilization of these links and not only the bi-directional links improves the routing protocol performance.

9-Security: Due to the nature of transmission medium, MANET routing protocol is vulnerable to many forms of attacks. They are more prone to security replay transmission, do spoofing threats than other general wired networks because the network structure is not strictly defined. Hello flood attack are common .Also a number of nodes keep on getting added as well as deleted from the network making it very easy for a malicious node to enter a network. Then it will be relatively easy for that node to snoop on network traffic, redirect traffic and flood the entire network. Security is very important to stop any kind of disruption of the network.

10-Power conservation: The nodes in the MANETs network can be laptops and constraint clients such as PDA's that are limited in battery power and therefore uses some standby mode to save the power. It is therefore very important that the routing protocol has support for these sleep modes.

11-Multiple routes: To reduce the number of reactions to topological changes and congestion multiple routes can be used. If one route becomes invalid, it is possible that another stored route could still be valid and thus saving the routing protocol from initiating another route discovery procedure.

12-Quality of Service Support: Most of the group communication technologies support real-time multimedia applications such as video conferencing, video streaming and distributed gaming. These applications require quality-of-service (QoS) aware multicast routing protocol to deliver the same data stream to a predefined group of receivers. Some sort of QoS is necessary to incorporate into the routing protocol.

13-Scalability: Routing protocols should be able to scale with the network size. Scalability can be broadly defined as whether the network is able to provide an acceptable level of service even in the presence of a large number of nodes. In MANET when the network size increases No. of packets send by node also increases that leads to drainage of limited battery power and network life time gets reduced thus scalability is major challenging issue.

14-Energy consumption: In MANETs networks each node participating in the network acts both router and a host and is willing to transfer packets to other nodes. For this purpose, a routing protocol should minimize control traffic. The concept of power as one of the deciding factor in route selection can be crucial in route discovery and route repair phase.

2.1 Characteristics of an Ideal Routing Protocol for MANETs Network

A routing protocol should have the following essential characteristics:

- It must be fully distributed.
- Adaptive to frequent change in topology
- Transmission should be reliable to reduce message loss.
- The convergence must be quick, once the network of the topology becomes stable.
- Optimal use of bandwidth, computing power, memory and battery power.
- It must provide a certain level of quality of service (QoS).
- Loop free ,Least control overhead ,Energy –aware , location-aware ,

The main design criteria for the routing protocols in MANETs are as follows:

- Scalability and Reliability
- Dynamic topology
- Maintenance
- Distributed and lightweight
- Simplicity and ease of implementation
- Fault Tolerance

III. CLASSIFICATION MANETs ROUTING PROTOCOLS

Routing is a core problem in networks for sending data from one node to another. Routing is point of research focus since the invention of commercialized mobile ad-hoc networks. Several MANET routing protocols have been designed for accurate, fast, reliable, scalable, stable, fairness, robust, QoS aware and energy efficient routing protocols for a high volume of changeable network topology. Such protocols must deal with the typical limitations of changeable network topology, which include high power consumption, low bandwidth, and high error rates. Till to date, number of different routing protocols for mobile ad-hoc networks have been proposed. To establish communication path between nodes, efficient routing protocols are needed. There are number of routing protocols currently available in MANETs. There is a need for a general technique to classify available protocols. As shown in Fig (1-a) MANETs routing protocols it can be divided into two general approaches first one depending on routing strategy and second one based on network structure .Fig(1-b) shown the classification according to the routing strategy, here the routing protocols can be categorized as table-driven and source initiated. Fig(1-c) shown the classification depending on the network structure, these are classified as flat routing, hierarchical routing, geographical (location based routing), power-aware routing and multicast routing [5].Here in this paper we will focus on routing protocols based on network structure because has gain greet interest from the researchers and industry .

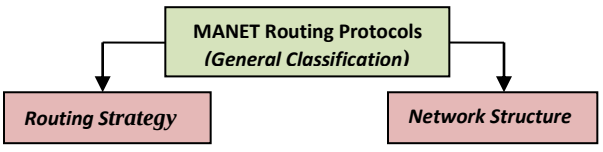


Fig (1-a) MANETs Routing Protocols Main Classification

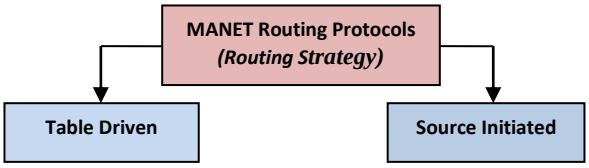


Fig (1-b) MANETs Routing Protocols Classification according to the routing strategy

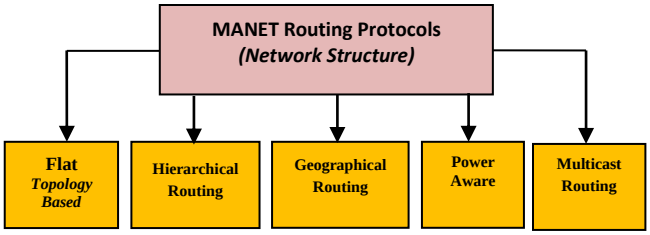


Fig (1-c) MANETs Routing Protocols Classification according to the Network Structure

3.1 Flat Routing (Uniform) or (Topology Based)

Topology based routing protocols depend on current topology of the network and cope with the dynamic nature of MANET. The topology-based routing protocols have limited performance when we are comparing with geographical (position based) routing protocols which uses additional information in order to determine the node location. Topology Based Routing schemes generally require additional node topology information during the routing decision process. Topology based routing can be further subdivided into proactive routing protocols(table-driven), reactive routing (on-demand) protocols, and hybrid routing protocols[6, 7]. Fig (2) shows the detailed taxonomy of Flat Routing (Topology Based).

3.1.1 Proactive or (Table Driven)

The network is under constant survey in order to know all possible routes between nodes at any given time; this means that routes are constantly being discovered, even if routes have not been invalidated .Maintain table for each node which contains the latest information of routes to nodes, to know its local neighborhood. This control messages are periodically exchanged. Examples of proactive protocols are Destination-Sequenced Distance Vector (DSDV), Optimized Link-State Routing (OLSR), Topology-Based Reverse Path Forwarding (TBRPF) Protocols, and Core-Extraction Distributed Ad hoc Routing (CEDAR). We can distinguish three types of the proactive routing protocols according to the algorithm based on each one.

a) Distance Vector routing: Distant vector protocol is also known as Distributed Bellman-Ford or RIP (Routing Information Protocol) .In a distance vector routing protocol, every host maintains a routing table containing the distances from itself to possible destinations or in other words contains all available destinations details, the next node to reach to destination and the number of hops to reach the destination[8]. Each routing table entry contains two parts: the next hop to the destination, and the distance to the destination. The distance metric might be the number of hops, the delay, the quality of links along the path, etc. The chosen next hops lead to the shortest path to the destination[9]. Using a distance vector protocol, the router simply forwards the packet to the neighboring host (or destination) with the available shortest path in the routing table and assumes that the receiving router will know how to forward the packet beyond that point[10].

Destination Sequenced Distance Vector (DSDV) Protocol

Is a proactive, hop-by-hop distance-vector routing protocol based on the classical Bellman–Ford routing algorithm proposed by (Charles Perkins and Bhagwatt 1994). It is a distributed, self-organized, and loop-free routing protocol suitable for dynamic networks[11]. Each node maintains a routing table that contains routing entries for all nodes in the network, and periodically advertises and broadcast routing updates of their routing information to their neighbors. Each entry in the routing table contains the destination node's address, next-hop node's address, the number of hops to reach, and the sequence number originated by the destination node. Nodes can forward packets to next-hops, and so on to the destination according to their routing tables. The sequence number is used to distinguish stale routes from new ones and thus avoid loop formation. The stations periodically transmit their routing tables to their immediate neighbors. A station also transmits its routing table if a significant change has occurred in its table from the last update sent. So, the update is both time-driven and event-driven. The routing table updates can be sent in two ways: a "full dump" or an "incremental" update[12].

b) Link State routing:

In link state protocols, a router doesn't provide the information about the destination instead it provides the information about the topology of the network. This usually consist of the network segments and links that are attached to that particular router along with the state of the link i.e., whether the link is in active state or the inactive state. This information is flooded throughout the network and then every router in the network then builds its own picture of the current state of all the links in the network[10].

Optimized Link State Routing (OLSR) Protocol

(Jacquet et al 1998) (OLSR) protocol is a proactive routing protocol where the routes are always immediately available when needed. OLSR is an optimization version of a pure link state protocol in which the topological changes cause the flooding of the topological information to all available hosts in the network. OLSR may optimize the reactivity to topological changes by reducing the maximum time interval for periodic control message transmission. Furthermore, as OLSR continuously maintains routes to all destinations in the network, the protocol is beneficial for traffic patterns where a large subset of nodes are communicating with another large subset of nodes, and where the [source, destination] pairs are changing over time[13].

c) QoS proactive routing protocols:

The primary goal of the QoS-aware routing protocols is to determine a path from a source to the destination that satisfies the needs of the desired QoS. The QoS-aware path is determined within the constraints of bandwidth, minimal search, distance, and traffic conditions. Since path selection is based on the desired QoS, the routing protocol can be termed QoS-aware. In the literature, numerous routing protocols have been proposed for finding QoS paths[14].The most popular examples of the proactive QoS routing protocols is CEDAR , QOLSR, DQRA , In the following section one of these QoS routing protocols is described.

Core-Extraction Distributed Ad hoc Routing (CEDAR) Protocol:

(Sivakumar et al 1999) have proposed the Core-Extraction Distributed Ad hoc Routing (CEDAR) algorithm. Is proactive protocol proposed for QoS routing in ad hoc networks for small to medium sized ad hoc networks, the network size may be tens to hundreds of nodes.In CEDAR the bandwidth is used as the only QoS parameter for routing. Most of the multimedia applications require the communication to meet stringent requirements on delay, delay-jitter, cost and other QoS metrics[15]. In these circumstances, the trend is to move from single constrained routing to multi constrained routing. The main function of multi-constrained QoS routing is to find a feasible path that satisfies multiple constraints simultaneously, which is a big challenge for MANETs where the topology may vary constantly.

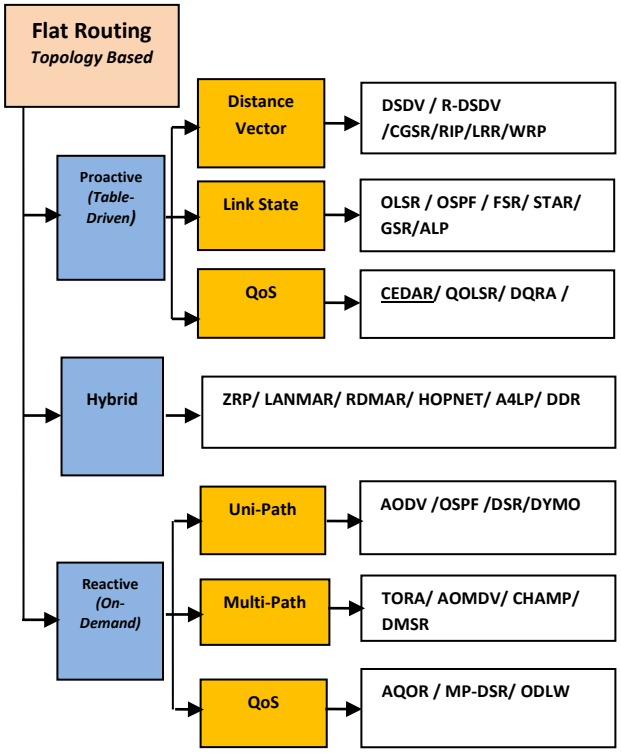


Fig (2) Flat Routing (Topology Based) Classification with subdivisions

CEDAR has three key components: (a) the establishment and maintenance of a self-organizing routing infrastructure, called the "core extraction". During the core extraction a set of nodes is dynamically elected to form the core of the network for performing route computations, (b) the propagation of the link-state of stable high-bandwidth links in the core through “increase/decrease" waves. During the link state propagation the bandwidth availability information of stable high bandwidth links is propagated to core nodes, while information regarding low bandwidth and unstable links is kept local and (c) a QoS route computation algorithm that is executed at the core nodes using only locally available state[16]. During the last phase, the core path is used to establish a route from the source to the destination[17]. In the CEDAR approach, the core provides an efficient low-overhead infrastructure to perform routing, while the state propagation mechanism ensures availability of link state information at the core nodes without incurring high overheads[18]. The greatest novelties of this technique were the core broadcast and link capacity dissemination mechanisms. These ensure efficient use of network resources and relatively accurate and up-to-date knowledge of the QoS state, where it is required. Furthermore, this protocol does not rely on a TDMA network. However, the problem of estimating available link capacities was left open[19].Table (1) shown a comprehensive comparison between there Flat Routing -Proactive protocols DSDV , OLSR and CEDAR.

3.1.2 Reactive Routing Protocols (On Demand)

This type of protocols attempts to establish routes between nodes only when they are needed or when routes are no longer valid. Thus, reactive routing protocols such as AODV try to establish a route to a destination only when needed. Discover routes only on demand basis and do not take initiative for finding a route. Reactive routing needs less memory and storage capacity than proactive protocols. They do not update route tables constantly. Reactive protocol is also known. as the on-demand routing protocol which they don’t maintain the routing information or activity of routing at the network nodes which there is no communication. In this protocol, the node wants to send a packet to the other node while for the route in an on-demand routing manner and establish the connection of route to transmit and receive the packet in order to manner. A packets will forward throughout the network by the flooding of the route occurs by the route discovery. The examples of reactive or on-demand routing protocols are: AODV, DYMO, TORA, AQOR, ARA [20].

a) Uni-Path routing: Even if several equally good paths are available, the unipath routing protocols use only one path at a time to a given destination. Such protocols as AODV, OSPF, DSR, and DYMO operate with this strategy. Most routing protocols are unipath or have a unipath mode of operation. A commonly used routing protocol called Open Shortest Path First (OSPF) operates in a unipath mode equal-cost multipath routing is enabled or turned on in the protocol[21].Route discovery and route maintenance are the two steps followed by each protocol. Route discovery: Source node first finds a route or several routes to the destination, when it needs to send packets to a destination. This process is called route discovery. Route maintenance: The source transmits packets with the route. The route may be broken during the transmission of packets, because the node on the route move away or go down. The broken route will be reconstructed. The process of detecting route breakage and rebuilding the route is called route maintenance.

Ad hoc On Demand Distance Vector (AODV)

AODV (Charles Perkins and Elizabeth Royer 1999; Charles Perkins and Elizabeth Royer 2002; Perkins et al., 2003; Chaudhry et al., 2005; Gorka Hernando et al 2009) is reactive routing protocol belongs to the unipath routing protocols; is on demand routing protocol, whenever a route from source to destination is required then only it develops a route .AODV created with the combination of Dynamic source routing (DSR) and Destination Sequenced Distance-Vector (DSDV); AODV use properties of route request (RREQ) and also route maintenance procedure from DSR and some features like sequence number, periodic updates, hop by hop count from DSDV routing protocol, every node knows its neighbors and the costs to reach them. AODV protocol of MANET doesn’t have a fixed topology in a network. This is basically needed for wireless communication for the nodes and

Table (1) Comparison between the Proactive protocols DSDV, OLSR and CEDAR

Protocol **Properties**	**DSDV** **Destination sequenced distance vector**	**OLSR** **Optimized Link State Routing Protocol**	**CEDAR** **Core-Extraction Distributed Ad Hoc Routing**
Proactive type	**Distance Vector**	**Link State- unicast**	**QoS Routing**
Based algorithm	Bellman-Ford algorithm	- Multipoint Relays (MPRs) - Optimized link state	- (CEDAR) algorithm
Advantages	- Distributed, self-organized, guarantees loop-free paths to each destination - Suitable routing protocol for dynamic networks - Less delay is involved in the route setup because of the availability of routes to all destinations at all times - DSDV maintains only the best path so the amount of space in routing table is reduced - avoid traffic with incremental updates	- Does not need central administrative system to handle its routing process - Well suited for an application which does not allow long delays in the transmission of data packets. - Improve the transmission quality	- It performs both routing and QoS path computation very efficiently with the help of core nodes - utilization of core nodes reduces the traffic overhead - Multi constrained routing - Find a feasible path that satisfies multiple constraints simultaneously - Network size may be tens to hundreds of nodes
Disadvantages	- Heavy control overhead because of the updates due to broken links - It is not suitable for highly dynamic networks - DSDV is inefficient due to the requirement of periodic update transmissions - To continue an up-to-date view of the network topology at all the nodes the updates are propagated throughout the network	- Needs more time to re-discovering the broken link. - Requires more power when discovering alternative route. - Is not feasible for highly dynamic networks because of the significant state propagation overhead when the network topology changes.	- The movement of the core nodes affects the performance of the protocol. –The update and gathering information of the whole network state at each core nodes could cause a significant or very high of control overhead. - does not require high maintenance overhead even for highly dynamic networks
Limitations	- DSDV doesn't support Multi path Routing. - It is difficult to determine a time delay for the advertisement of routes - It is difficult to maintain the routing table's advertisement for larger network. - The route is decided through the sequence number -	- Wider delay distribution - This protocol needs that each node periodically sends the updated topology information throughout the entire network, this increase the protocols bandwidth usage - Needs more time to re-discovering the broken link. - OLSR requires more power when discovering alternative route.	- Route establishment and computation is relied on core nodes. - Estimating available link capacities was left open - Unicast-based "core broadcast" for reliability
Enhanced versions (Extensions)	- Research work on improvement of DSDV is still active. Many improved protocols based on DSDV have been developed. These improvements of DSDV include Global State Routing (GSR), Fisheye State Routing (FSR), Ad Hoc On-Demand Distance Vector Routing (AODV)	-HOLSR and EE-OLSR proposed based on OLSR with hierarchical architecture and Energy Efficient -QOLSR (Support quality of service) -Geo-OLSR(support geographical location information service)	- MCEDAR (Multicast) provides the robustness of mesh based routing *protocols* - Recent research works towards combines CEDAR with DSR and AODV to propose DSRCEDAR and AODVCEDAR
Performance Metrics and QoS parameters	**Throughput**: - Least very low when compared with DSR, OLSR and AODV	- High when compared with other link state protocols[22]	- Stable high bandwidth -MCEDAR is the enhanced version support multicasting mechanism
	Packet/dropped/loss: - high	- Packet loss rate is less because most of the Packets sent and received is among the MPR nodes	
	End to end delay and jitter: - Least and remains constant as the number of nodes increase in the networks.	- Average end to end delay and least compared to DSDV	
	Packet Delivery Ratio (PDR): - Increases initially then Low compared to OLSR	- higher packet delivery ratio with compared to DSDV	
	Routing Overhead: - Very high for a slight increase in the number of nodes	- Low (reduces the routing overhead)	
	Caching Overhead: - Medium	- High	-low

path from which request packet come. Following information is contained in the packet header for route request:

- Source node IP address
- Broadcast ID
- Current sequence number for the destination

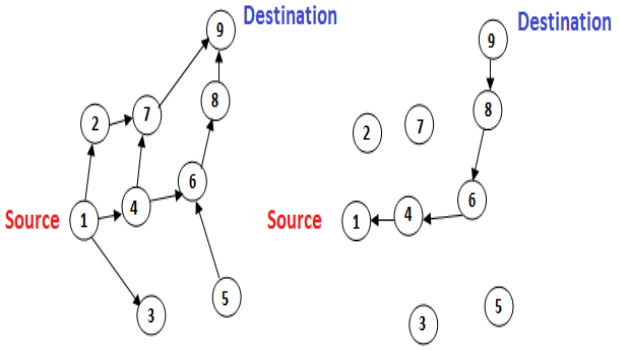


Figure (3): AODV Route Discovery

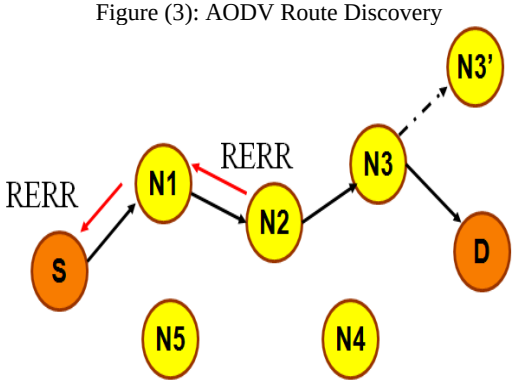


Figure (4) AODV Route Maintenance Process

During a route discovery process as shown in Figure (3), the source node broadcasts a route query packet to its neighbors. If any of the neighbors has a route to the destination, it replies to the query with a route reply packet; otherwise, the neighbors rebroadcast the route query packet. Finally, some query packets reach to the destination. The route maintenance process in AODV is very simple (as shown in Figure (4)). When the link in the communication path between node S (source node) and node D (Destination node) breaks the upstream node that is affected by the break, in this case node N2 generates and broadcasts a RERR message. The RERR message eventually ends up in source node S. After receiving the RERR message, node S will generate a new RREQ message.

b) Multi-Path routing:

In single path routing protocols, a single route is discovered between source and destination. Discovery of multiple routes between source and destination in single route discovery is done by multipath routing. Multipath Routing is the process of distributing the data from source node to destination node over multiple paths. Multipath algorithms permit traffic multiplexing over multiple paths. Multipath Routing performs better by proper usage of network resources. Multipath routing protocols provides better throughput and reliability than single path protocols. The main goals of multipath routing protocols are to maintain reliable communication, to reduce routing overhead by use of secondary paths, to ensure load balancing, to improve quality of service, to avoid the additional route discovery overhead[23].

Temporally Ordered Routing Algorithm (TORA)

TORA (Park and Corson 1997) is a source-initiated on-demand routing protocol, which uses a link reversal algorithm and provides loop-free multipath routes to a destination node. In TORA, each node maintains its one-hop local topology information and also has the capability to detect partitions. TORA is proposed to operate in a highly dynamic mobile networking environment. The key design concept of TORA is the location of control messages sent to a very small set of nodes near the occurrence of a topological change. The protocol performs three basic functions: (1) route creation, (2) route maintenance, and (3) route erasure. During the route creation and maintenance phases, the nodes use a height metric, which establishes a direct acyclic graph (DAG) rooted at the destination. Therefore, links are assigned a direction (upstream or downstream) based on the relative height metric of neighboring nodes, as shown in Figure (5). The process for establishing a DAG is similar to the query/reply process in lightweight mobile routing. In times of node mobility, the DAG route is broken, and route maintenance is necessary to reestablish a DAG rooted at the same destination. Timing is an important factor for TORA because the height metric depends on the logical time of link failure. TORA assumes all nodes have synchronized clocks. In TORA, there is a potential for oscillations to occur, especially when multiple sets of coordinating nodes are concurrently detecting partitions, erasing routes, and building new routes based on each other. Because TORA uses inter nodal coordination, its instability problem is similar to the "count-to-infinity" problems.

QoS reactive routing protocols:

In any given network, there are two types of flows in general: BE (Best Effort) flows which require the data to be reliably delivered to the destination, and QoS flows such as RT (Real Time) which apart from reliability, requires some additional constraints such as available bandwidth, delay, etc. to be satisfied[24] . Reusing BE routing methods for QoS-aware routing is not feasible since BE routing performs these tasks based on a single measure, usually hop-count while QoS-aware routing, however, must take into account multiple QoS measures and requirements. This section discusses different QoS-aware routings in MANETs from different perspectives including its challenges, classifications, algorithms and comparisons.

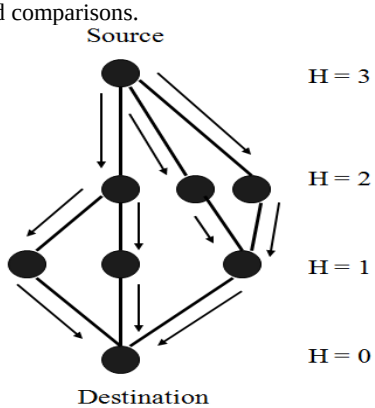


Figure (5) TORA routing scheme

Ad hoc QoS on-demand routing (AQOR)

Is a resource reservation and signaling algorithm proposed by (Xue and Ganz 2003). AQOR is a reactive QoS routing protocol guarantee smallest end-to-end delay and bandwidth in MANETs [18]. Uses limited flooding to discover the best route available in terms of smallest end-to-end delay with bandwidth guarantee. A route request packet includes both bandwidth and end to-end delay constraints .This protocol mainly used to provides end-to-end QoS support in terms of bandwidth and end-to-end delays in MANETs. It is a resource reservation-based routing and signaling scheme that allows AQOR to make admission and resource reservation decisions. AQOR integrates on-demand route discovery between the source and the destination, signaling functions for resource reservation and maintenance and hop-by-hop routing. AQOR is also a source initiated, on-demand routing protocol. It is built upon AODV routing, performing exploration of routes only when required. The route discovery mechanism is in on-demand mode, broadcasting the RREQ and RREP packets between the source and destination nodes. The route discovery mechanism starts when the node broadcasts RREQ packets with QoS requirements to its neighboring nodes. The neighboring nodes that satisfy the requirement add a route entry to the source node's routing table and forward the RREQ until it reaches the destination. When the RREQ reaches the destination node, an RREP is sent back along the reverse route, reserving bandwidth at each node. Once the source node receives the RREP, it starts sending data out along the reserved route. AQOR uses timers to detect route breaks and to trigger route recovery. If any node fails to receive a data packet before its reservation expires, a route recovery mechanism is triggered. Source node starts the route discovery process all over again by broadcasting an RREQ packet. Initiating a route discovery process each time a route break occurs can lead to high end-to-end delays. AQOR uses routing tables for keeping track of its routes. Every time a route failure occurs, AQOR must update its routing table entries, which may sometimes result in inconsistent entries due to the high dynamic nature of the network topology. To avoid possible loops during route exploration, AQOR uses a route sequence number to indicate the freshness of the control packets for each follow. The sequence number is maintained at each mobile node aware of the follow. The initial sequence number of any follow is 0. When sending out a route control packet for a follow (e.g., RREQ, REP, or RERR), the initial node will increase its current sequence number by 1 and attach the value to the packet .In AQOR, the route with the shortest end-to-end delay, given it satisfies the bandwidth requirement, is selected. This protocol mainly used to bandwidth and End-to-End delay. This protocol is guarantee for smallest end - to- end delay and bandwidth[18]. A Bandwidth and end-to-end delay constraints are included in route request packet. Table (2) shown a comparison between three types of Flat-routing (Reactive Protocols) AODV, TORA and AQOR.

3.1.3 Hybrid Routing Protocols: is a combination of reactive and proactive routing protocols. These protocols have the advantage of both proactive and reactive routing protocols to balance the delay which was the disadvantage of table driven protocols and control overhead (in terms of control packages). Main feature of Hybrid Routing protocol is that the routing is proactive for short distances and reactive for long distances E.g. ZRP, LANMAR, HOPNET, and DDR.

Zone Routing Protocol (ZRP)

(Pearlman and Haas 1999), as seen, proactive routing uses excess bandwidth to maintain routing information, while reactive routing involves long route request delays. Reactive routing also inefficiently floods the entire network for route determination[25]. ZRP divides the entire network into overlapping zones of variable size where routing inside the zone is performed using proactive approach and outside the zone is performed using reactive approach[26].

3.2 Hierarchical routing

The idea behind hierarchical routing is to divide the hosts of self-organized networks into a number of overlapping or disjointed clusters[27] .Hierarchical-Network is used when the size of network inside a MANET increases tremendously[7]. Hierarchical routing protocols organize the network as a tree of clusters, where the roles and functions of nodes are different at various levels of the hierarchy. Routes are constructed according to the node's position in the virtual hierarchy[11]. Non-uniform hierarchical routing protocols can be further sorted into three subcategories: zone-based, cluster-based, and core-based. These protocols are categorized according to the organization of the mobile nodes, their respective management and their routing functions. Fig (6) shown the subcategories of MANETs hierarchical routing protocols.

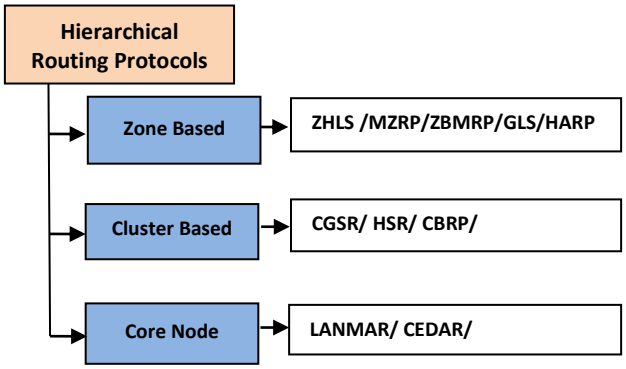


Fig (6) Hierarchical Routing Classification

3.2.1 Zone Based

With zone-based hybrid routing algorithm technique each node has a local scope and different routing strategies are used, inside and outside the scope, as communications pass across the overlapping scopes. Given this flexibility, a more efficient overall routing performance can be achieved. Compared to maintaining routing information for all nodes in the whole network, mobile nodes in the same zone know how to reach each other with a smaller cost. In some zone-based routing protocols, specific nodes act as gateway nodes and carry out inter-zone communication. Therefore, the network will contain partitions or a number of zones. The Zone Routing Protocol (ZRP) is a MANET zone-based hierarchical routing protocol[28].

Zone-Based Hierarchical Link State Routing Protocol (ZHLS)

ZHLS (Joa Ng and Lu 1999) is a hierarchical routing protocol, and it is a zone-based hierarchical protocol that makes use of location information in a novel peer-to-peer hierarchical routing approach[27]. In ZHLS, the network is divided into non-overlapping clusters (zones) without any masters (zone-heads) as

shown in Fig. (7). In ZHLS, mobile nodes are assumed to know their physical locations with assistance from a locating system like GPS. In ZHLS protocol, the network is divided into non-over lapping zones as in cellular networks[29]. Each node has own node ID and a zone ID, which is calculated by using the GPS[20]. This topology made up to two levels: node level topology and zone level topology. Each node knows the node connectivity within its own zone and the zone connectivity information of the entire network. The link state routing is performed by employing two levels: node level and global zone level. ZHLS does not have any cluster head in the network like other hierarchical routing protocols.

3.2.2 Cluster Based

A cluster based routing protocol is the most popular hierarchical routing technique .The process of dividing the network into interconnected sub structures is called clustering and the interconnected substructures are called clusters. The cluster head (CH) of each cluster act as a coordinator within the substructure. Each CH acts as a temporary base station within its zone or cluster. It also communicates with other CHs[30]. The Cluster based routing provides an answer to address nodes heterogeneity, and to limit the amount of routing information that propagates inside the network. A cluster-based routing protocol uses specific clustering algorithms for cluster
head election. Mobile nodes are grouped into clusters and cluster heads take the responsibility for membership management and routing functions. CGSR will be introduced in this section as an example of cluster-based mobile ad hoc network routing protocols. Some cluster-based MANETs routing protocols potentially support a multilevel cluster structure, such as hierarchical state routing (HSR) [27].

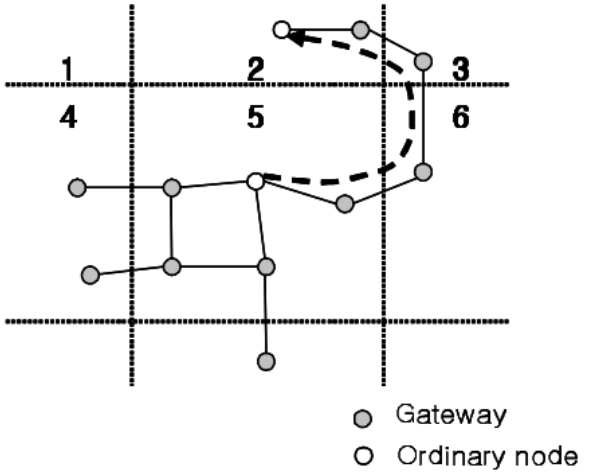


Fig (7) ZSLS routing(Zone 5 and 6 are connected but 2 and 5 are not) [31]

Cluster-Head Gateway Switch Routing (CGSR)

(CGSR) (Chiang et al 1997) [32] is a hierarchical routing protocol. The cluster structure improves performance of the routing protocol because it provides effective membership and traffic management. Besides routing information collection, update and distribution, cluster construction and cluster-head selection algorithms are important components of cluster based routing protocols. The CGSR protocol is a clustering scheme that uses a distributed algorithm called the Least Cluster Change (LCC)[33]. CGSR extends DSDV with a cluster framework concept that increases protocol scalability, also, heuristic methods like priority token scheduling, gateway code scheduling, and path reservation are used to improve the protocol's performance. On the other hand, setting up structure in a highly dynamic environment can adversely affect protocol performance since the structure might not persist for a very long time [34]. CGSR is multi-channel routing protocol is generally used in TDMA or CDMA-based networks. They combine channel assignment and routing functionality. CGSR is a non-uniform hierarchical protocol, which is based to forming clusters among nodes and selecting a cluster head to control routing to outside the cluster area.[35]. By aggregating nodes into clusters controlled by cluster heads, a framework for developing additional features for channel access, bandwidth allocation, and routing is created. Nodes communicate with the cluster head, which, in turn, communicate with other cluster heads within the network. Selecting a cluster head is a very important task because frequently changing cluster heads will have an adverse effect on the resource allocation algorithms that depend on it. Thus cluster stability is of primary importance in this scheme. The LCC algorithm is stable in that a cluster head will change only under two conditions: when two cluster heads come within the range of each other or when a node gets disconnected from any other cluster. CGSR is an effective way for channel allocation within different clusters by enhancing spatial reuse. The explicit requirement of CGSR on the link layer and MAC scheme is as follows: (1) Each cluster is defined with unique CDMA code and hence each cluster is required to utilize spatial reuse of codes. (2) Within each cluster, TDMA is used with token passing. Gateway nodes are defined as those nodes which are members of more than one cluster and therefore need to be communicating using different CDMA codes based on their respective cluster heads. The main factors affecting routing in these networks are token passing (in cluster heads) and code scheduling (in gateways). This uses a sequence number scheme as in DSDV to reduce stale routing table entries and gain loop-free routes. A packet is routed through a collection of these cluster heads and gateways in this protocol. In CGSR, when forwarding a packet, a node firstly checks both its cluster member table and routing table and tries to find the nearest cluster-head along the
routing path. As shown in Figure (8), when sending a packet, the source (node 1) transmits the packet to its cluster-head (node 2). From the cluster-head node 2, the packet is sent to the gateway node (node 3) that connecting to this cluster-head and the next cluster-head (node 5) along the route to the destination (node 8). The gateway node (node 6) sends the packet to the next cluster-head (node 7), i.e. the destination cluster-head. The destination cluster-head (node 7) then transmits the packet to the destination (node 8).

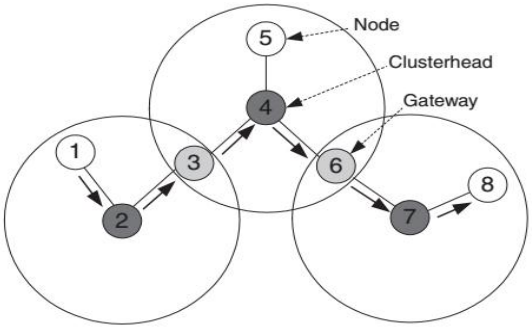


Fig (8) Cluster Structure in CGSR [32]

Table (2) Comparison between the reactive protocols AODV, TORA, AQOR

Protocol / Properties	AODV		TORA	AQOR
Reactive type	**Uni-path** (unicast)		**Multi-path**	**QoS**
Advantages	- Provides loop-free multipath routes to a destination node - Reliable and offers quick adaptation to dynamic link conditions - Less delay in connection setup. - Low processing and memory overhead - Low network utilization - Determines unicast between sources and destinations. - The HELLO messages supporting the route maintenance are range limited, so they do not cause unnecessary overhead in the network - It uses destination sequence numbers to ensure loop freedom at all times (even in the face of anomalous delivery of routing control messages) - Solving problems (such as ``counting to infinity'') associated with classical distance vector protocols. - Coping up with dynamic topology and broken links - Highly Scalable		- Supports multiple routes between source and destination. Hence, failure or removal of any of the nodes quickly resolved without source intervention by switching to an alternate route to improve congestion. - Provides the supports of link status sensing and neighbor delivery, reliable, in-order control packet delivery and security authentication. - TORA discovers new route faster and more effectively to the destination when the old route is broken as it invokes route repair mechanism locally also high route cache hit ratio in TORA. - Operate in a highly dynamic mobile networking environment	- Provide QoS support in terms of bandwidth and end-to-end delay. - AQOR's route firm is quick and reliable. - AQOR's traffic measurements and admission decisions are accurate and provide high channel utilization. - The route with the least delay is chosen by the source
Disadvantages	- High route discovery latency, AODV does not discover a route until a flow is initiated - Multiple Route Reply packets in response to a single Route Request packet can lead to heavy control overhead - Routing table entries are purged (deleted) after a certain period of time even if any or some of the links are valid. - Periodic exchange of beacons for detecting broken links consumes bandwidth		- Same as on-demand routing protocols. - Not much used since DSR and AODV outperform TORA. - TORA support unidirectional links and multiple routing paths	- Initiating a route discovery process each time a route break occurs can lead to high end-to-end delays. -
Enhanced versions	- AOMDV(multicast) , PAAODV(power-aware) , EAODV(Energy-aware)		- PDTORA	-
Limitations	- Requirement on broadcast medium - AODV lacks support for high throughput routing metrics - It is vulnerable to misuse - AODV doesn't support multiple routing paths - Large delay caused by route discovery process		- Not scalable by any means - TORA assumes all nodes have synchronized clocks - In TORA, oscillations may occur when coordinating nodes currently execute the same operation	- Every time a route failure occurs, AQOR must update its routing table entries, which may sometimes result in inconsistent entries due to the high dynamic nature of the network topology
Based algorithm	- Builds on the DSDV algorithm		- Link reversal algorithm and Directed Acyclic Graph (DAG)	- Resource reservation-*based* routing and signaling *algorithm*
Performance Metrics and QoS parameters	**Throughput**	- Poor for more than 20 Mobiles	- Better throughput	- Based on Bandwidth and Delay only - **Bandwidth Utilization:** Minimum bandwidth - **end to end delay:** Maximum end-to-end delay
	Packet dropped	- Minimum	- Moderate	
	Jitter and end to end delay	- high initially in AODV but after some time it is very low - as the number of nodes are increasing the delay of AODV is increasing	- High compared to DSR	
	Packet Delivery Ratio(PDR):	- High	- High	
	Routing Overhead	- AODV has less traffic overhead , but compared to DSDV is high	- Low compared to DSR	
	Bandwidth Utilization	- Overhead on the bandwidth because RREQ travels from node to node in the process of discovering the route info on demand, it automatically sets up the reverse path from all nodes back to the source	- Since TORA does not require a periodic update, consequently communication overhead and bandwidth utilization is minimized.	
	Caching Overhead	- Low	- Medium	

3.2.3-Core Node

In core node-based routing protocols, critical nodes are dynamically selected to compose a "backbone" for the network. The "backbone" nodes carry out special functions, such as the construction of routing paths and propagation of control/data packets. Optimized Link State Routing (OLSR) and Core Extraction Distributed Ad hoc Routing (CEDAR) protocols are typical core node-based MANET routing protocols[28].

Landmark Ad Hoc Routing (LANMAR)

In the Fisheye State Routing protocol (FSR) (Pei et al 2000a), every node in the network needs to maintain whole network topology information. This strictly limits its scalability. The Landmark Ad hoc Routing (LANMAR) (Pei et al 2000b) is proposed as a modification of FSR and aims to gain better scalability. In contrast to FSR, LANMAR belongs to the non-uniform routing category of mobile ad hoc networks. In LANMAR, mobile nodes are divided into predefined logical subnets according to their mobility patterns, i.e., all nodes in a subnet are prone to move as a group. A landmark node is pre-specified for every logic subset to keep track of the subnet. Using LANMAR; every mobile node has a hierarchical address that includes its subnet identifier. A node maintains the topology information of its neighbors and all landmark nodes, which represent logical subnets. Similar to FSR, neighboring nodes in LANMAR periodically exchange topology information and the distance vector of landmark nodes. When a source sends packets to the destination inside its neighboring scope (i.e., the source and the destination belong to the same subnet), desired routing information can be found from the source's routing table. Otherwise, the subnet identified in the destination node's address will be searched. Then, according to the distance vector, the packets will be routed towards the landmark node of the logical subset. Compared to FSR, LANMAR is more efficient because the need to exchange topology information is reduced substantially. However, LANMAR assumes that nodes are grouped into subsets according to their movement patterns and the membership of each subnet remains unchanged during the lifetime of the network, so it is only suitable for specific application scenarios. Table (3) illustrates a comparison between a three major Hierarchical protocols ZHLS, CGSR and LANMAR.

3.3Geographic Position (Information Assisted)

Geographic routing (also called geo-routing or position-based routing) is a routing principle that relies on geographic position information. Geographical routing protocols are topological independent, developed for large and distributed network operations. Generally, in traditional MANETs, the nodes are addressed only with their IP addresses[29]. But, in case of location-aware routing mechanisms, the nodes are often aware of their exact physical locations in the three-dimensional world within the network. The proactive zones act as collectors of the GPSs, which are embedded in nodes, are used to update information in tables in position-based algorithms. That makes position-based algorithms different from the table driven and on demand algorithms[36].This type of protocols is mainly proposed for mobile wireless networks and based on the idea that the source sends a message to the geographic location of the destination instead of using the network address. Geographic routing is a technique to deliver a message to a node in a network over multiple hops by means of position information[5]. Node use broadcast to know location of one hope neighbor. Due to this position -based routing required minimal routing overhead and also avoid delay and latency due to localized forwarding. Routing decisions are not based on network addresses and routing tables; instead, messages are routed towards a destination location. Geographic Position (Information assisted) protocols is used to eliminate the limitations of topology based routing by using additional information. It gives the better performance in dynamic topologies because the packets are forwarded to its destination with respect to its position. Each node determines its own position and for determining the position of the network node the different positioning schemes are used such as GPS, GPRS etc. Location-aware routing does not require the routes establishment and maintenance. No routing information is stored. Typically, a node selects the next hop for packets forwarding by using the physical position of its one-hop neighbors and the physical position of the destination node; positioning information of the networks' nodes are usually obtained via queries offered through some location service[34] .Fig (9) shows the different categories of the location aware or geographical position routing technique. Generally this type can subdivided into three main approaches with respect to path strategy and packet forwarding : greedy (single-path), flooding (multi-path) and hierarchical [37, 38]. In the next section we will discuss and explain each category aided with examples of protocols such as GPRS, DREAM and LAR.

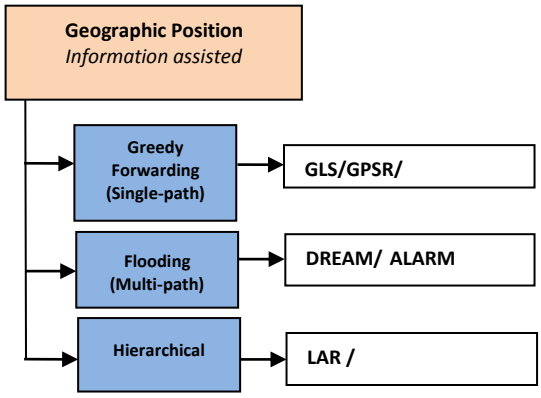


Fig (9) Geographic Position Routing Classification

3.3.1Greedy Forwarding (Single-path)

The shortest path route is an example of a single-path strategy, where one copy of the message is in the network at any time. Most single-path strategies rely on two techniques: greedy forwarding and face routing. Greedy forwarding tries to bring the message closer to the destination in each step using only local information. Greedy forwarding is used when the message is able to advance from source towards the destination (Figure 10-a).
It does not imply route establishment or maintenance and the next hop[38].

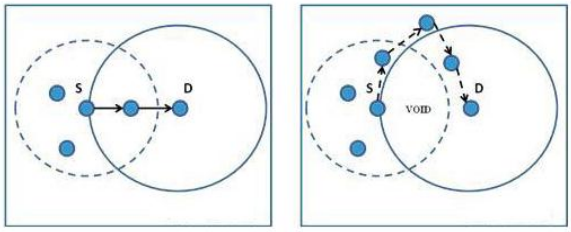


Figure (10) (a) Greedy (b) Greedy Failure at node S

Table (3) Comparison between the Hierarchical protocols ZHLS, CGSR and LANMAR

Protocol / Properties	ZHLS	CGSR	LANMAR
Hierarchical type	**Zone Based**	**Cluster Based**	**Core Node**
Advantages	- The network is geographically divided into non-overlapping zones - ZHLS assumes that each node has a location system such as GPS and the geographical information is well known - ZHLS defines two levels of topologies node level and zone level. - Single point of failure and traffic bottlenecks can be avoided	- Multi-channel routing protocol is generally used in TDMA or CDMA-based networks - Non-uniform hierarchical protocol - Forming clusters among nodes and selecting a cluster head to control routing to outside the cluster area. - Uses a sequence number scheme to reduce stale routing table entries and gain loop-free routes. - Simpler addressing scheme	- Distributive , adaptive , hierarchical routing - Robust in rapid topological change - Mobile nodes are divided into predefined logical subnets according to their mobility patterns - Guarantees the shortest path from a source to a destination if the destination is located within the scope of the source.
Disadvantages	- ZHLS does not have any cluster head in the network like other hierarchical routing protocols In ZHLS is that all nodes must have a preprogrammed static zone map in order to function or in other words Static zone map is required	- Highly dynamic environment can adversely affect protocol performance - Frequently changing cluster heads will have an adverse effect on the resource allocation algorithms that depend on it. - Since additional time is required to perform cluster head reselection, time to recover from link failure is higher than DSDV and WRP	- LANMAR assumes that nodes are grouped into subsets according to their movement patterns and the membership of each subnet remains unchanged during the lifetime of the network, so it is only suitable for specific application scenarios - Assumption of group mobility. Nodes may not have the best route to distant destinations
Routing Table/Overhead	- Large communication overhead in the network, because In ZHLS all network nodes construct two routing tables, an intra-zone routing table and an inter-zone routing table	- Reduces the size of the routing table as well as the size of routing update messages. - Since each node only maintains routes to its cluster head in CGSR, routing overhead is lower than compared to DSDV and WRP.	- Reduces both routing table size and control overhead for large MANETs.
Limitations	- ZHLS take time to search new route when route is disconnected because it search only one route. In particular, real-time application is severely-impacted by this delay.	- Both cluster member and routing tables need to be updated - Uses DSDV as an underlying routing scheme	- LANMAR is only suitable for specific mobile applications
Based algorithm	ZHLS algorithim	- Clustering algorithm based on the lowest identifier or the highest connectivity - Least Cluster Change (LCC)	- Scoped routing algorithm (e.g., FSR) - Binding algorithm
Cluster structure	- No masters Multiple gateways between clusters	- Single gateway - clustering algorithm based on the lowest identifier or the highest connectivity	- Group mobility is assumed so that relative relationship among - mobile nodes in a group doesn't change over time and it results in a natural clustering.
Scalability	- Support high scalability	- High scalable.	- Improves routing scalability for large MANETs with the assumption that nodes under a landmark move in groups.

Forwarding

Thus, each node forwards the message to the neighbor that is most suitable from a local point of view. The most suitable neighbor can be the one who minimizes the distance to the destination in each step (Greedy). The decision is made according to the optimization criteria of the algorithm and does not guarantee that a packet reaches its destination. Metrics can be hop count, geographic distance, progress to destination, direction, power, cost, delay, a combination of these, etc. If the message has reached a node which has no closer neighbors to the destination (a void or hole), a recovery procedure is necessary (Figure 10-b) making the forwarding method a hybrid. Recovery from such a concave node can be done through flooding or perimeter (face) forwarding.

Greedy Perimeter Stateless Routing protocol (GPSR)

(Karp and Kung 2000) proposed GPSR routing protocol which uses the location of node to selectively forward the packets on the basis of distance. The packets are forwarded on a greedy basis by selecting the node closest to the destination. The best path was also calculated through a node which was farther in geometric distance from the destination. This process continues until the destination is reached. In some cases the best path may be through a node which is farther in distance from the destination node. In such scenario right hand rule is applied to forward around the obstacle and resume the greedy forwarding as soon as possible[39, 40].

3.3.2 Flooding-based (Multi-path)

In flooding-based approaches, messages are flooded through the whole network area or portion of the area. A simple flooding geocast algorithm works as follows: A node broadcasts a received packet to all neighbors as shown in Fig(11) provided that this packet was not already received before in order to avoid loops and endless flooding. A node delivers a packet if the own location is within the specified destination region, which is included in each geocast packet. This is a simple and robust but not efficient approach, since location information is not used for forwarding in order to reduce the number of packets[41].

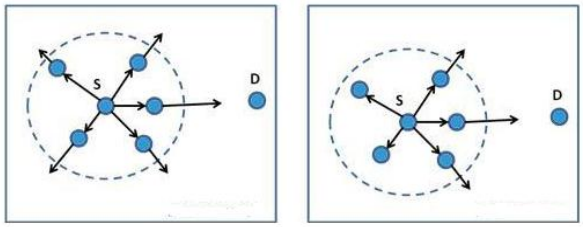


Figure (11) (a) Flooding (b) Restricted Flooding

Distance Routing Effect Algorithm for Mobility (DREAM)

DREAM (Basing et al., 1998) provides location service for position-based routing. In this framework, each node maintains a position database that stores position information about other nodes in the network. It can therefore be classified as an all-for-all approach, which means all nodes work as the location service providers, and each node contains all other nodes location information. An entry in the position database includes a node identifier, the direction and distance to the node, together with the time-stamp of entry creation. Each node regularly floods packets to update the position information maintained by the other nodes. Since the accuracy of the position information in the database depends on its age, a node can control the accuracy of its position information available to other nodes by adjusting the frequency of sending.

3.3.3 Hierarchical Approaches

The third forwarding strategy is to form a hierarchy in order to scale to a large number of mobile nodes. Some strategies combine the nodes location and hierarchical network structures by using the zone based routing. Others use the dominating set routing. Some others present a two level hierarchy within them; if the destination is close to the sender (in number of hops), packets will be routed based on a proactive distance vector. Greedy routing is used in long distance routing[39].

Location Aided Routing (LAR):

(Ko and Vaidya 1998, Ljubica Blazevie et al 2005) presents the LAR protocol which utilizes location information to minimize the search space for route discovery towards the destination node [39]. LAR is a reactive unicast routing protocol which based on DSR. LAR aims to reduce the routing overhead for the route discovery and it uses the Global Positioning System (GPS) to obtain the location information of a node. LAR essentially describes how location information such as GPS can be used to reduce the routing overhead in an ad hoc network and ensure maximum connectivity. Location-Aided Routing is an example of restricted directional flooding routing protocols; however, partial flooding is used in LAR for path discovery purpose. Hence, LAR proposes the use of position information to enhance the route discovery phase of reactive Ad-Hoc routing approaches. Table (4) contain a comparison between a three geographic Position protocols GPRS, DREAM and LAR.

3.4 Power Aware routing protocols

As MANET lack of fixed infrastructure and mobiles nodes in MANET are battery driven, in such environment energy efficiency is an important consideration to increase the network[42] .Since the nodes in MANETs are mobile, the routing and power management become critical issue[43] .Several power aware routing schemes have been proposed for MANETs networks Fig(12). The main objective of power aware routing scheme is to minimize the power consumption and maximize the network lifetime. The network lifetime is defined up to the moment when a node runs out of its own battery power for the first time[44].This classification of protocols is based on the consumption of energy during transmission. I.e., Energy required to transmit a signal is approximately proportional to dX, where d is distance and is the attenuation factor or path loss exponent, which depends on transmission medium. When X=2, which is optimal case, Transmitting a signal half of the distance required one fourth of energy and if anode is in the middle, will spend another fourth of energy for the second half. Thus data will transmit for half of the energy than through direct transmission.

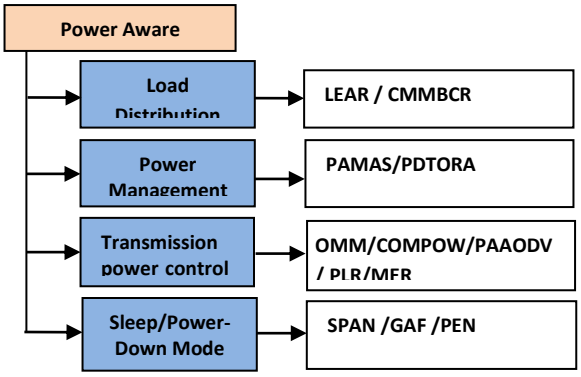


Figure (12) Power and Energy-aware subdivisions

Table (4) Comparison of Geo-protocols GPSR ,DREAM and LAR

Protocol / Attribute	GPSR	DREAM	LAR
Geographic Position Type	**Greedy Forwarding**	**Flooding**	**Hierarchical**
Advantages	-Guarantees a good PDR especially in the high density of nodes, -Data forwarding overhead is low -Increasing of efficiency when more node added to the network - local maxima can be found easily	- Packet loss is higher than GPSR - No delay for routing discovery methods	-Localized route discover - Restricted directional flooding -Reduces routing overhead -Reduce the number of nodes to which the route request is propagated.
Disadvantages	-impossible to find the optimal path -Scalability occurs when increasing in network diameter and mobility -Delay increases at high mobility, -generates a large number of control packets for high speeds	- Requires GPS. - Flooding, can influence the performance of the basic algorithm. -A recovery method is necessary when the destination node is not in the given direction.	-Based on source routing, flooding is used if no location information is available -Low performance when various optimization techniques are not implemented like alternative definition of request zone, another adaptation of request zone, and so on
Energy consumption	Low	High	High
Routing metric	Closest distance	Shortest Path	Shortest Path
Packet Delivery Ratio(PDR),	High	Low	Low
Jitter and end to end delay	Lower delay	Long delay	Long delay
Routing/Communication Overhead	- High	-minimizes routing overhead	reduces overhead by limiting search to requested zone only, reduce traffic, no need of hello message

3.4.1 Load Distribution

This approach is also based on active communication energy. The main goal of this approach is to balance the amount of energy usage among all mobile nodes and to maximize the lifetime of network by avoiding over-utilized nodes when selecting a routing path. The protocol selecting underutilized nodes rather than the shortest route. This may result in longer routes, but packets are routed only through energy rich intermediate nodes. Protocols based on this approach do not only provide route with lowest energy, but prevent certain nodes from being overloaded, and thus increases the network lifetime. The most popular examples of this approach is Localized Energy-Aware Routing (LEAR), Conditional Max-Min Battery Capacity Routing protocols (CMMBCR).

Localized Energy-Aware Routing (LEAR)

The LEAR (Woo et al. 2001) protocol directly controls the energy consumption. In particular, it achieves balanced energy consumption among all participating mobile nodes. The LEAR protocol is based on DSR, where the route discovery requires flooding of route-request messages. When a routing path is searched, each mobile node relies on local information of remaining battery level to decide whether or not to participate in the selection process of a routing path. An energy-hungry node can conserve its battery power by not forwarding data packets on behalf of others. Decision -making process in LEAR is distributed to all relevant nodes, and the destination node does not need wait or block it order to find the most energy efficient path. Upon receiving a route -request message, each mobile node has the choice to determine whether or not to accept and forward the route -request message depending on its remaining battery power (E_r).When it is higher than a threshold value (Th_r), the route-request message is forwarded; otherwise, the message is dropped. The destination will receive a route-request message only when all intermediate nodes along the route have good battery levels. Thus, the first arriving message is considered to follow an energy-efficient as well as a reasonably short path.

3.4.2 Power Management

Power management technique is used to reduce the energy consumed in the MANETs interface of battery powered mobile devices. The design of best possible power management policies needs to explicitly account for the dissimilar performance requirements posed by different application scenarios such as latency, throughput and other performance metrics[43]. Power management techniques have been studied comprehensively in the context of CPU, memory and disk management in the past. The main idea is to switch devices to the low-power state in periods of inactivity. As compared with traditional techniques in operating systems, power management in communication devices requires distributed coordination between two (or multiple) communicating entities, as all the entities have to be in the active mode for a successful communication.

Power Aware Multi-Access (PAMAS) routing protocol :PAMAS is an extension to the AODV protocol. It uses a new routing cost model to discourage the use of nodes running low on battery power. The lifetime of the network is improved significantly. This routing protocol saves energy by turning off radios when the nodes are not in use. Although, it was implemented on the AODV protocol, the technique used is very standard and can be used with any on-demand protocol. The energy aware protocol works only in the routing layer. Advantage of PAMAS protocol is that this protocol saves 40-70% of battery power by intelligently turning off radios when they cannot transmit or cannot receive packets. This protocol tends to increase the throughput of the network as compared to other power aware routing protocols .One of the disadvantages of PAMAS protocol is broadcasting problem. In this protocol, a broadcast may collide with another transmission at some receiver[42].

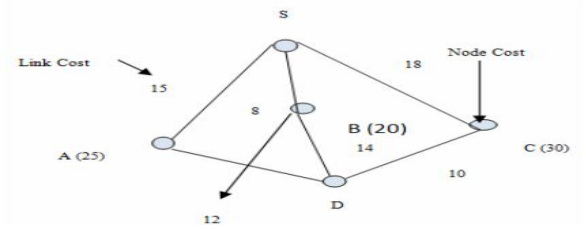
3.4.3 Transmission power control

Transmission power control approach can be achieved with the help of topology control of a MANET [23]. The transmission power determines the range over which the signal can be coherently received, and is therefore crucial in determining the performance of the network (throughput, delay, and power consumption) [24]. Power aware routing protocols based on transmission power control finds the best route that minimizes the total transmission power between a source and destination. It is equivalent to a graph optimization problem, where each link is weighted with the link cost corresponding to the required transmission power. Finding the most power efficient (min-power) route from source to destination is

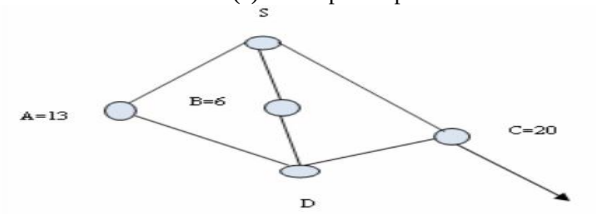
equivalent to finding the least cost path in the weighted graph. A routing algorithm essentially involves finding an optimal route on a given network graph where a vertex represents a mobile node and an edge represents a wireless link between two end nodes that are within each other's radio transmission range. In this paper, we reviewed various power aware routing protocols explained each one of them by taking our own examples and also introduced the new power aware routing protocol i.e. PADSR.

Online Max-Min Routing Protocol (OMM)

Online Max-Min (OMM) power-aware routing protocol for MANETs networks dispersed over large geographical areas to support applications where the message sequence is not known. This protocol uses two different metrics of the nodes in the network to optimizes the lifetime of the network as well as the lifetime of individual nodes by maximizing the minimal residual power (max-min), which helps to prevent the occurrence of overloaded nodes, the other metric is Minimizing power consumption (min-power) . In most applications that involve MANETs, power management is a real challenge and can be done at two complementary levels (1) during communication and (2) during idle time. The OMM protocol maximizes the lifetime of the network without knowing the data generation rate in advance. The metrics developed showed that OMM had a good empirical competitive ratio to the optimal online algorithm that knows the message sequence and the max-min achieves over 80% of the optimal node lifetime for most instances and over 90% of the optimal node lifetime for many problem instances. OMM protocol uses Dijkstra's algorithm to find the optimal path between source-destination pair. This min-power path consumes the minimal power (P min).In order to optimize the second metric, the OMM protocol obtains multiple near-optimal min-power paths that do not deviate much from the optimal value (i.e., less than $zP\min$ where $z \geq 1$) and selects the best path that optimizes the max-min metric. Figure-(13) shows an example of the algorithm for a given source and destination pair. In Figure-3 (a), S → *B* → *D* is the min-power path as it consumes the minimal energy (P min=22) i.e. path cost is 22. If z = 2, alternative paths S→ *A* → *D* (path cost=27) and S → *C* → *D* (path cost=28) can be considered since their path costs are within the tolerance range ($zP\min = 44$)[45, 46].



(a) Min-power path



(b) Max-min path

Figure (13) Min-power path and max-min path in the OMM protocol [44]

3.4.4 Sleep/Power-Down Mode

This approach is based on saving the energy during inactivity means when node is idle. Nodes can save the energy during inactivity by switching into sleep/power-down mode when there is no data to transmit or receive. This leads to considerable energy savings, especially when the network environment is characterized with low duty cycle of communication activities. However, it requires well-designed routing protocol to guarantee data delivery even if most of the nodes sleep and do not forward packets for other nodes.

SPAN protocol

SPAN protocol is a power saving mechanism that reduces power consumption of nodes by retaining the capacity and coordinating with the underlying MAC layer[47]. SPAN protocol operates between the routing layer and the MAC layer. SPAN coordinates the "stay-awake and sleep" cycle of the nodes and also performs multi-hop packet routing within the ad hoc network. While other nodes remain in the power saving mode and periodically check if they should remain awaken and become a coordinator. SPAN adaptively elects coordinators by allowing each node to use a random back-off delay to decide whether to become a coordinator in the network and rotates them in time. This technique not only preserves network connectivity, it also preserves capacity, decreases latency and provides significant energy saving. Other advantage of the SPAN protocol is that the master nodes play an important role in routing by providing a routing backbone and control traffic as well as channel contention is reduced. Disadvantage of SPAN protocol is that the amount of power saving increases slightly as density decreases.

3.5 Multicast Routing Protocols

Multicast is the delivery of information to a group of destinations simultaneously, using the most efficient strategy to deliver the messages over each link of the network only once, creating copies only when the links to the destinations split. Multicast routing protocols for MANET use both multicast and unicast for data transmission[29].In recent years, a number of multicast protocols for ad hoc networks have been proposed. Based on the routing structure, they can broadly be classified into two categories: tree-based protocols and mesh-based protocols [48].Fig (14) below shown the basic subdivisions of the Multicast routing protocols .

3.5.1 Tree Based Multicast Routing Protocols: In tree-based protocols, there exists a single path between any sender-receiver pair. Tree-based protocols have the advantage of high multicast efficiency (which is defined as the ratio of the total number of data packets received by all receivers to the total number of data packets transmitted or retransmitted by senders or intermediate nodes). However, tree-based protocols are not robust against frequent topology changes and the packet delivery ratio (which is defined as the ratio of the number of data packets delivered to all receivers to the number of data packets supposed to be received by all receivers) drops at high mobility.

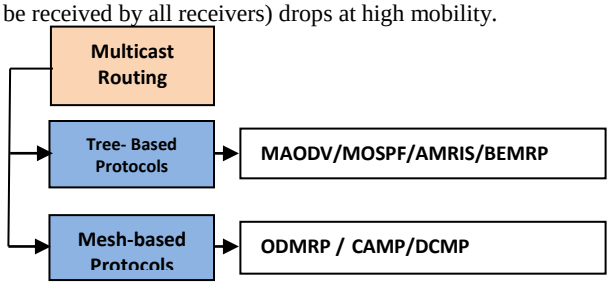


Figure (14) Multicast Routing Protocols subdivisions based on the Network structure

Table (5) Comparison among different MANETs routing approach[50]

Protocol	Structure /Route Computation	#Routes	Stored Information	Update Period	Update information	Update Dest.
Flat Routing (Topology Based) ---- Reactive (On-Demand)						
AODV	Flat –Reactive/ broadcast QUERY	Multiple	Next hops for desired dest.	Event-driven RM	ROUTE-ERROR	Source
TORA	Flat-Reactive/ broadcast QUERY	Multiple(DAG)	Neighbors' heights	Event-driven	Node's height	Neighbors
AQOR	Flat-Reactive/ Limited QUERY	Multiple	bandwidth and end to-end delay	Periodical	bandwidth ,end to-end delay, signaling	Neighbors
DSR	Flat-Reactive/ broadcast QUERY	Multiple	Routes to desired Dest.	Event-driven RM	ROUTE-ERROR	Source
Flat Routing (Topology Based) --- Proactive (Table Driven)						
OLSR	Flat-Proactive/ distributed	Multiple	MPR nodes, link load, delay, bandwidth	Periodical	HELLO and TC messages	Neighbors
DSDV	Flat-Proactive/ distributed	Single	Distance vector	Hybrid	Distance vector	Neighbors
FSR	Flat-Proactive/ distributed	Single or multiple	Entire topology , closer nodes	Periodicals(dif. freq.)	Link state of fisheye scope ,distant nodes	Neighbors
WRP	Flat-Proactive/ distributed	Single	Dist./routing/link-cost table, MRL	Hybrid	Dist. Vec., List of Resp.	Neighbors
GSR	Flat-Proactive/ distributed	Single or multiple	Entire topology	Periodical	All nodes link state	Neighbors
	Flat- Routing --- Hybrid routing Protocols					
ZRP	Flat-Hybrid Proactive(intra)/Reactive(inter)	Single or multiple	Local (within zone), topology	Periodical	Link state of nodes in the zone	Neighbors
ZHLS	Hierarchy-Hybrid/ Proactive/Reactive (hier. addr.)	Multiple	Local(inter zone) , (intra zone) topology	Period./Event-driven	Node/Zone , link state	Zone/all nodes
Hierarchical protocols						
CSGR	Hierarchy-cluster-Proactive/ distributed	Single	Cluster. Member. Table, Dist.Vec.	Periodical	Cluster. Member. Table, Dist.Vec.	Neigh.&Clus. head
LANMAR	Hierarchy-Core node	Single or multiple	Entire topology, dist.Vec. of landmark nodes.	Periodical	landmark distance vector, Next Hop Address, sender's LMDV	Next Hop /Destination
CEDAR	Hierarchy-Proactive/ core broadcast QUERY	Single	Core/other nodes: global/local	Period./Event-driven	Dynamic/stable Link state	Neigh./Core nodes
Geographic Position (Information assisted) Routing Protocols						
GPSR	Geographic- Greedy Forwarding/	Single	Greedy mode(provides all nodes with their neighbors' positions)	Periodic beaconing	Perimeter mode (Beacon to the broadcast MAC address, IP, position.)	greedily /neighbors
DREAM	Geographic-flooding / distributed	Multiple	position info., about other nodes in the network	Periodical	distances separating nodes	*Dest.,* node, neighbor node.
LAR	Geographic-reactive	multiple	Location of the destination, distance to define the requested zone.	periodic update of beacons	expected zone of the destination	request zone
Power-Aware , Energy –Efficiency routing Protocols						
LEAR	Power-aware	multiple	Battery Power (E_r) and threshold value (Th_r)	Periodical	energy efficient, shortest path	*Destination*
PAMAS	Power-aware/ RTS-CTS message exchange	multiple	Await Packet, Idle ,wait CTS, BEB	Periodical	CTS or Busy Tone	Neighbors
SPAN	Power-aware/ broadcast QUERY	multiple	routing backbone and control traffic	Periodical		Neighbors
OMM	Power-aware/ Min-power , and max-min	multiple	Link cost, Node cost, tolerance range, graph optimization algorithm	Periodical	zP_{min} , P_{min}, link costs	Dest., Neighbors
Multicast Routing Protocols						
MAODV	Multicast	Multiple	requesting node's ,next hop	periodically	RREP , forward path	next hop
ODMRP	Multicast	Multiple	source ID ,sequence number in its message cache	periodically	group membership and multicast routes	Neighbors

Multicast Ad Hoc On-Demand Distance Vector Routing Protocol (MAODV)

(Elizabeth Royer and Perkins 1999) proposed MAODV routing protocol [49] witch discovers multicast routes on demand using a broadcast route-discovery mechanism. A mobile node originates a Route Request (RREQ) message when it wishes to join a multicast group, or when it has data to send to a multicast group but it does not have a route to that group. Only a member of the desired multicast group may respond to a join RREQ. If the RREQ is not a join request, any node with a fresh enough route (based on group sequence number) to the multicast group may respond. If an intermediate node receives a join RREQ for a multicast group of which it is not a member, or if it receives a RREQ and it does not have a route to that group, it rebroadcasts the RREQ to its neighbors. As the RREQ is broadcast across the network, nodes set up pointers to establish the reverse route in their route tables. A node receiving a RREQ first updates its route table to record the sequence number and the next hop information for the source node. This reverse route entry may later be used to relay a response back to the source. For join RREQs, an additional entry is added to the multicast route table. This entry is not activated unless the route is selected to be part of the multicast tree. If a node receives a join RREQ for a multicast group, it may reply if it is a member for the multicast group's tree and its recorded sequence number for the multicast group is at least as great as that contained in the RREQ. The responding node updates its route and multicast route tables by placing the requesting node's next hop information in the tables, and then unicasts a Request Response (RREP) back to the source node. As nodes along the path to the source node receive the RREP, they add both a route table and a multicast route table entry for the node from which they received the RREP, thereby creating the forward path, see Figure 15.

3.5.2 Mesh-based protocols provide redundant routes for maintaining connectivity to group members. The low alleviated due to redundant routes. Mesh-based protocols are robust to node mobility. However, redundant routes cause low multicast efficiency.

On-demand Multicast Routing Protocol (ODMRP)

ODMRP (On-demand Multicast Routing Protocol) [51]is mesh based, and uses a forwarding group concept (only a subset of nodes forwards the multicast packets). A soft-state approach is taken in ODMRP to maintain multicast group members. No explicit control message is required to leave the group. In ODMRP, group membership and multicast routes are established and updated by the source on demand. When a multicast source has packets to send, but no route to the multicast group, it broadcasts a Join-Query control packet to the entire network. This Join-Query packet is periodically broadcast to refresh the membership information and update routes, see Figure 2.

When an intermediate node receives the Join-Query packet, it stores the source ID and the sequence number in its message cache to detect any potential duplicates. The routing table is updated with the appropriate node ID (i.e. backward learning) from which the message was received for the reverse path back to the source node. If the message is not a duplicate and the Time-To-Live (TTL) is greater than zero, it is rebroadcast. When the Join-Query packet reaches a multicast receiver, it creates and broadcasts a "Join Reply" to its neighbors. When a node receives a Join Reply, it checks if the next hop node ID of one of the entries matches its own ID. If it does, the node realizes that it is on the path to the source and thus is part of the forwarding group and sets the FG_FLAG (Forwarding Group Flag).It then broadcasts its own Join Table built upon matched entries. The next hop node ID field is filled by extracting information from its routing table. In this way, each forward group member propagates the Join Reply until it reaches the multicast source via the selected path (shortest). This whole process constructs (or updates) the routes from sources to receivers and builds a mesh of nodes, the forwarding group.

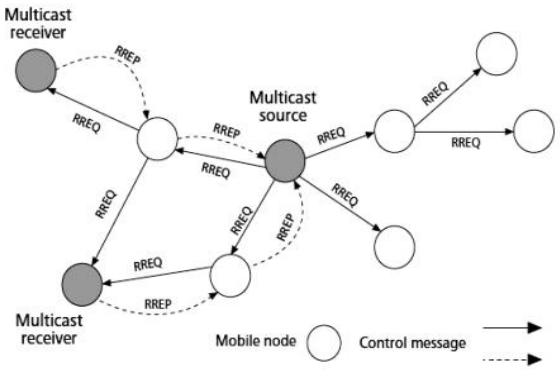


Fig (15) MOADV route discovery mechanism[52]

IV.CONCLUSION AND FUTURE WORKS

This paper presents a comprehensive survey and overview of the recent routing process and strategies of MANETs. In the literature most of the authors classified MANETs routing protocols into: proactive, reactive and hybrid. However this classification is not covering other important routing mechanisms in MANETs such as Hierarchical Routing (HR), Geographical Routing (GR), Power Aware (PA) and Multicast Routing (MR). In this paper, several existing routing protocols for MANETs networks were described. One example for each category of routing strategy was discussed. While this survey we found that AODV, OLSR and TORA from the Flat–routing approach are powerful, highly adaptive, efficient and scalable distributed routing algorithm. Those protocols are efficient and adaptable for different application specifically real time applications such as video streaming or video conferencing. From the Hierarchical approach we found that ZHLS, LANMAR and CGSR are highly scalable and in the other hand has least communication overhead, means that these protocols are capable for delay sensitive applications and also compatible for tactical scenarios where the nodes are spread out a huge coverage area. The Geographic routing protocols such as (GPSR, DREAM and LAR) scale better for MANETs mainly for two reasons: 1) there is no necessity to keep routing tables up-to-date and 2) no need to have a global view of the network topology and its changes. The power aware and multicast routing provides a low communication overhead and it benefits in large scale MANETs and it will be feasible for some applications such as VANET, iMANET, smart cities and IoT.

REFERENCES

[1] M. K. Nazir, Rehman, R.U. and Nazir, A., "A Novel Review on Security and Routing Protocols in MANET," Communications and Network, vol. 8, pp. 205 - 218, 2016.

[2] S. R. Inamdar and R. M. Yadahalli, "Paradigm Shift in Routing Approaches for High Speed MANET Applications," European Journal of Engineering Research and Science, vol. 2, 2017.

[3] D. S. R. M.Vijayalakshmi, "EFFICIENT MULTICAST ROUTING PROTOCOL FOR MANETS,"

International Journal of Computer Networks and Wireless Communications (IJCNWC),, vol. 6, pp. 56 - 61, 2016.

[4] Bharathi, "A Review and Comparative Study on Various Routing Protocols for MANET," International Journal of Innovative Research in Science,Engineering and Technology, vol. 5, pp. 20790 - 20795, 2016.

[5] R. Suma and B. Premasudha, "Geographical routing protocols for mobile ad hoc networks-a survey on their performance analysis," International Journal of Research in Engineering and Technology eISSN, pp. 2319-1163, 2013.

[6] B. A. Mahmood and D. Manivannan, "Position based and hybrid routing protocols for mobile ad hoc networks: a survey," Wireless Personal Communications, vol. 83, pp. 1009-1033, 2015.

[7] D. Muhit, "Performance Evaluation Of Aodv, Dsr, Dsdv Mobile Ad-Hoc Protocols On Different Scenarios: An Analytical Review," International Journal of Advances in Computing and Information Technology, 2012.

[8] B. Jagdale, P. Patil, P. Lahane, and D. Javale, "Analysis and Comparison of Distance Vector, DSDV and AODV Protocol of MANET," International Journal of Distributed and Parallel Systems, vol. 3, p. 121, 2012.

[9] Y. Lu, W. Wang, Y. Zhong, and B. Bhargava, "Study of distance vector routing protocols for mobile ad hoc networks," in Pervasive Computing and Communications, 2003.(PerCom 2003). Proceedings of the First IEEE International Conference on, 2003, pp. 187-194.

[10] K. Gorantala, "Routing protocols in mobile ad-hoc networks," A Master'thesis in computer science, pp-1-36, 2006.

[11] A. Boukerche, Handbook of algorithms for wireless networking and mobile computing: CRC Press, 2005.

[12] S. Mohapatra and P. Kanungo, "Performance analysis of AODV, DSR, OLSR and DSDV routing protocols using NS2 Simulator," Procedia Engineering, vol. 30, pp. 69-76, 2012.

[13] S. Ade and P. Tijare, "Performance comparison of AODV, DSDV, OLSR and DSR routing protocols in mobile ad hoc networks," International Journal of Information Technology and Knowledge Management, vol. 2, pp. 545-548, 2010.

[14] D. S. P. a. D. K. K. SANJEEV GANGWAR, "Mobile Ad Hoc Networks: A Comparative Study of QoS Routing Protocols " IJCSET, vol. 2, pp. 771-775, 2012.

[15] S. Gangwar and K. Kumar, "Mobile Ad Hoc Networks: A detailed Survey of QoS Routing Protocols," International Journal of Distributed and Parallel Systems (IJDPS) Vol, vol. 2, 2011.

[16] R. Sivakumar, P. Sinha, and V. Bharghavan, "CEDAR: a core-extraction distributed ad hoc routing algorithm," IEEE Journal on Selected Areas in Communications, vol. 17, pp. 1454-1465, 1999.

[17] T. Kilinkaridis, "Routing Protocols for Wireless Ad Hoc Networks."

[18] S. Raja and K. Alagarsamy, "ROUTING PROTOCOLS IN MANET QOS SURVEY," International Journal of Innovative Research in Advanced Engineering (IJIRAE), vol. 2, pp. 169 - 157, 2015.

[19] X. Wang, "Mobile Ad-Hoc Networks: Protocol Design," 2011.

[20] R. K. P. Behra Rajesh Umashankar, "A Comparative Study of Topology and Position Based Routing Protocols in Mobile Ad Hoc Networks," International Journal of Advanced Research in Computer Science & Technology (IJARCST), vol. 2, pp. 72 - 75, 2014.

[21] M. B. Banu and P. Periyasamy, "A Survey of Unipath Routing Protocols for Mobile Ad Hoc Networks," International Journal of Information Technology and Computer Science (IJITCS), vol. 6, p. 57, 2013.

[22] V. Deepalakshmi, T. Muthuramalingam, and M. R. Vijayalakshmi, "Current Survey of Routing Protocols in MANETS," in International Journal of Engineering Research and Technology, 2014.

[23] B. Kukreja and S. Kambhra, "Performance Comparison of Routing Protocols in MANET," International Journal of Computer Science and Network Security (IJCSNS), vol. 14, p. 108, 2014.

[24] N. Zhang and A. Anpalagan, "Comparative review of QoS-aware on-demand routing in ad hoc wireless networks," Wireless Sensor Network, vol. 2, p. 274, 2010.

[25] N. Beijar, "Zone routing protocol (ZRP)," Networking Laboratory, Helsinki University of Technology, Finland, pp. 1-12, 2002.

[26] S. Goyal, "ZONE ROUTING PROTOCOL (ZRP) IN AD-HOC NETWORKS," vol. 3, ed: IJREAS, 2013.

[27] S. K. Sarkar, T. Basavaraju, and C. Puttamadappa, Ad hoc mobile wireless networks: principles, protocols and applications: CRC Press, 2007.

[28] N. H. Saeed, M. F. Abbod, and H. S. Al-Raweshidy, "MANET routing protocols taxonomy," in Future Communication Networks (ICFCN), 2012 International Conference on, 2012, pp. 123-128.

[29] G. V. Kumar, Y. V. Reddyr, and D. M. Nagendra, "Current research work on routing protocols for MANET: a literature survey," international Journal on computer Science and Engineering, vol. 2, pp. 706-713, 2010.

[30] M. Rezaee and M. Yaghmaee, "Cluster based routing protocol for mobile ad hoc networks," INFOCOMP Journal of Computer Science, vol. 8, pp. 30-36, 2009.

[31] B. Lee, C. Yu, and S. Moh, "Issues in scalable clustered network architecture for mobile ad hoc networks," Handbook of Mobile Computing, pp. 1-27, 2004.

[32] C.-C. Chiang, H.-K. Wu, W. Liu, and M. Gerla, "Routing in clustered multihop, mobile wireless networks with fading channel," in proceedings of IEEE SICON, 1997, pp. 197-211.

[33] M. M. Artimy, W. Robertson, and W. J. Phillips, "Algorithms and Protocols for Wireless and Mobile Ad Hoc Networks," ed: John Wiley & Sons, 2009.

[34] M. C. STEFANO BASAGNI , SILVIA GIORDANO , IVAN STOJMENOVIC, Mobile ad hoc networking: John Wiley & Sons, 2004.

[35] D. Pramanik, "Classification of Ad Hoc Routing Protocols," International Journal of Innovative Technology & Adaptive Management (IJITAM) vol. 1, 2013.

[36] S. B. B. Rammyaa "A REVIEW OF POSITION BASED ROUTING PROTOCOLS FOR MOBILE AD

HOC NETWORKS," International Journal of Network and Mobile Technologies, vol. 2, pp. 77-81, 2011.

[37] I. Stojmenovic, "Position-based routing in ad hoc networks," IEEE communications magazine, vol. 40, pp. 128-134, 2002.

[38] A. M. Popescu, I. G. Tudorache, B. Peng, and A. H. Kemp, "Surveying position based routing protocols for wireless sensor and ad-hoc networks," International Journal of Communication Networks and Information Security, vol. 4, p. 41, 2012.

[39] S. Kaur and A. K. Gupta, "Position based routing in mobile Ad-hoc networks: An overview," IJCST, vol. 3, 2012.

[40] A. Gupta and S. D. Sharma, "A Survey on Location Based Routing Protocols in Mobile Ad-hoc Networks," Int. J. Comput. Sci. lnforrn. Technol, vol. 5, pp. 994-997, 2014.

[41] C. Maihofer, "A survey of geocast routing protocols," IEEE Communications Surveys & Tutorials, vol. 6, pp. 32-42, 2004.

[42] R. K. Yadav, D. Gupta, and R. Singh, "Literature Survey on energy efficient routing protocols for mobile ad hoc networks," International Journal of Innovations and Advancement in Computer Science, 2015.

[43] S. Kumara, E. S. Ranab, and E. A. Sharmac, "Energy Efficient Routing Techniques for Mobile Ad Hoc Networks," 2014.

[44] S. Mishra and B. K. Pattanayak, "Power Aware Routing In Mobile Ad Hoc Networks-A Survey," ARPN Journal of Engineering and Applied Sciences, vol. 8, pp. 173-189, 2013.

[45] S. Mishra and B. K. Pattanayak, "Power Aware Routing In Mobile Ad Hoc Networks-A Survey," ARPN Journal of Engineering and Applied Sciences, vol. 8, 2013.

[46] A. Kumaravel and M. Chandrasekaran, "A complete study on power aware routing protocol for mobile adhoc network," IOSR Journal of Electrical and Electronics Engineering, vol. 1, pp. 71-5, 2014.

[47] V. Bapuji, S. Sharma, R. L. Naik, D. Ramesh, and B. Manjula, "Quality of service for mobile Ad-hoc wireless networks," international Journal on computer Science and Engineering, vol. 3, pp. 1573-1577, 2011.

[48] C.-C. Huang and S.-C. Lo, "A comprehensive survey of multicast routing protocols for mobile ad hoc networks," 網際網路技術學刊, vol. 9, pp. 25-34, 2008.

[49] E. M. Royer and C. E. Perkins, "Multicast operation of the ad-hoc on-demand distance vector routing protocol," in Proceedings of the 5th annual ACM/IEEE international conference on Mobile computing and networking, 1999, pp. 207-218.

[50] X. Zou, B. Ramamurthy, and S. Magliveras, "Routing techniques in wireless ad hoc networks classification and comparison," in Proceedings of the Sixth World Multiconference on Systemics, Cybernetics and Informatics, SCI, 2002, pp. 1-6.

[51] Y. Yi, "On-demand multicast routing protocol (ODMRP) for ad hoc networks," draft-yi-manet-odmrp-00, 2003.

[52] M. R. BAKER and M. A. Akcayol, "A survey of multicast routing protocols in Ad-hoc networks," Gazi University Journal of Science, vol. 24, pp. 451-462, 2011.

International Journal of Computer Applications Technology and Research
Volume 6–Issue 3, 159-167, 2017, ISSN:-2319–8656

Minimizing the Impact of Forecast Error on Government Monetary and Fiscal Policy through Forecasting Software

Onu Fergus U. (Ph.D)
Department of Computer Science,
Ebonyi State University,
Abakaliki - Nigeria

Nnanna Emmanuel E.
Department of Computer Science,
Ebonyi State University,
Abakaliki - Nigeria

Abstract: Forecasting is an important part of governments' monetary and fiscal policies. Every forward-looking government uses economic forecasting as a necessary prerequisite for the success of its monetary and fiscal policies. Though economic forecasts are necessary, as they are essential underlying features of governments' monetary and fiscal policies, economic forecasting is a difficult task. The difficulty in economic forecasting arises from the interplay of variables which certainly produces forecast errors. This study focused on minimizing the impact of forecast error on government monetary and fiscal policies through the use of suitable forecasting software. Hence in this study, forecasting software are examined as necessary tools for minimizing the impact of forecast errors on governments' monetary and fiscal policies. Data is collected from both primary and secondary sources to elicit useful information from stakeholders in Forecasting Industry. The result of the study shows that there is a positive relationship between the use of forecasting software and minimizing the impact of forecast error on governments' monetary and fiscal policies. The study revealed that the use of sound software, devoid of mathematical errors and inappropriate methods, will minimize the impact of forecast error on governments' monetary and fiscal policies. The study also revealed that the major way to ensure an effective forecasting process and reduce forecast errors is to make use of intelligent forecast software that raises forecast accuracy.

Keywords: monetary and fiscal policy, economic forecasting, forecasters, forecastability, forecast error and forecast accuracy.

1 INTRODUCTION

There are no perfect forecasts; they always contain some error. Though it is a well known fact that there is no perfect forecast, it is nonetheless important to emphasize this fact at the onset. Forecasting is a process, and like any other process, it has inputs, outputs, objectives and owners. Forecasting is measurable and is applied to bring benefits to its owner(s). Better processes generally yield better results; analogously, better forecasting processes yield better forecasts. In the words of Diebold [1] "Forecasting is important - forecasts are constantly made in business, finance, economics, government, and many other fields, and they guide many important decisions. As with anything else, there are good and bad ways to forecast."

The danger of not having an effective forecasting process is that everyone is forecasting and every forecast is different. What every economy needs is one set of numbers with which every economic policy maker can operate. This set of numbers should be developed using an effective forecasting process. Forecast error is an inevitable occurrence in the forecast process. Factors like randomness, variation, uncertainty of forecast variables guarantees the inevitability of forecast error. What is important is to reduce the error, and to focus on the most effective process of reducing the error. Challenges in business forecasting, such as increasing accuracy and reducing bias, are best met through effective management of the forecasting process. Effective management, we believe, requires an understanding of the realities, limitations, and principles fundamental to the process. When management lacks a grasp of basic concepts like randomness, variation, uncertainty, and forecastability, the organization is apt to squander time and resources on expensive and unsuccessful fixes [2].

Economic forecasting as a prerequisite for a forward-looking macroeconomic policy has come to stay. Governments use economic forecasting to frame their monetary and fiscal policies. However, in producing and using economic forecasts, it is important for governments to be aware of the limitations of such forecasts mainly due to the impact of forecast errors. Such awareness will enable government to focus on the most effective process of reducing forecast error while choosing a given forecasting technique.

Therefore, the main objectives of this study among others are:

i. To identify the various techniques used in economic forecasting.

ii. To identify the causes of error in economic forecasting.

iii. To identify ways of improving forecast accuracy.

iv. To elucidate the ways of minimizing the impact of forecast error on government's monetary and fiscal policies.

This study in relevant in identifying the economic variabilities that limit forecast accuracy and in so doing, identifies ways of minimizing the impact of forecast errors on government's monetary and fiscal policies.

2. LITERATURE REVIEW

2.1 Forecasting (Definition and Evolution)

Forecasting is about predicting the future as accurately as possible given all the information available including historical data and knowledge of any future that might impact the forecasts [3]. Forecasting is estimating how the sequence of observations will continue into the future [3]. Economic forecasting is both a science and an art. It is a science in the sense that the rightful application of statistical tools will definitely improve forecasting accuracy. It is an art in the sense that empirical data seldom provide an unequivocal answer, so the user must choose between alternative relationships to select those equations that will provide the most accurate forecasts.

Forecasting is as old as man. The process of foresight has always been primary for humans. Blessed with the ability to perceive linear time, mankind has always sought to learn from the past and present to improve its lot in the future. Early forecasting by the primitive man was purely reactive or based on the assumption that whatever happened last period would pretty much happen again this period. Lacking reliable information on which to base decisions, the early man turned to drawing inferences from the natural world – identical patterns were identified, analyzed, and turned into actionable insights. This is called Naïve forecasting. Moving forward into the Renaissance and Industrial Age periods, man discovered statistical demand forecasting. This is quite a big leap forward. A time series of previous values could be used to create charts displaying aggregate demand one period after another. It projected forward what happened in the past few periods or simply added X% to the previous number. This discovery led to further advancement in forecasting - fitting curves through historical demand, creating moving averages, and trend lines. Even seasonality could be incorporated into the calculation. In the late twentieth century, predictive analytics were introduced. Demand planning helped by incorporating hierarchical and causal effects into the forecast. A forecast can statistically predict monthly or weekly demand patterns. The twenty-first century saw the emergence of Structured Quantified Forecast. This is a more complex process of forecasting. This process incorporates demand modeling and demand sensing. This system leverages more granular and downstream data to get a cleaner demand signal and reduce volatility and bullwhip effects. It includes techniques that are usually associated with short-term demand sensing to dramatically increase long-term accuracy and reduces the manual intervention to make things works. The most recent development in the forecast industry in this century is the introduction of Machine Learning. This is a form of artificial intelligence that captures and models complex patterns that shape demand signal, enabling forecasters to continuously fine tune the signal-to-noise ratio. This cutting edge technology identifies hidden patterns and trends that are extremely difficult and/or time consuming to uncover through other approaches, such as statistical analytics or human analysis. The results from this system are then used to refine future analysis, automatically making the system smarter and more accurate over time.

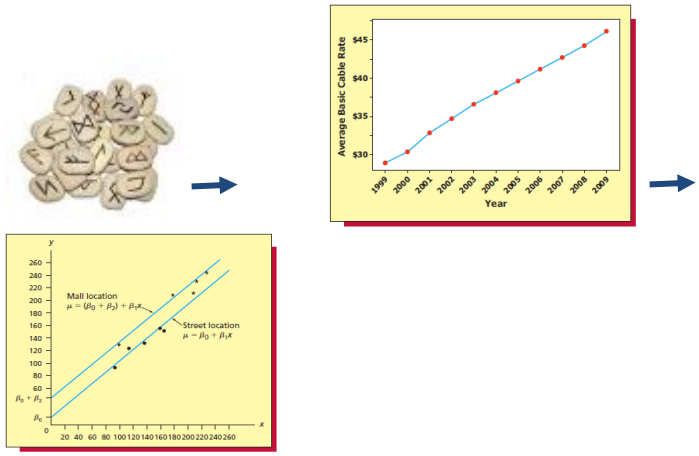


Figure 2.1. Evolution of forecasting from scrutinizing symbols through inventing statistics, predictive analytics to structured, quantified forecasts

2.2 Forecasting Techniques

All forecasting methodologies can be divided into two broad techniques: Qualitative Techniques and Quantitative Techniques. Qualitative techniques are based on expert judgments and opinions. Such techniques are usually employed when historical data on the variable being forecast are either not applied or unavailable. Judgment and opinion forecasts are largely intuitive, and historical analogy relies on comparisons. This technique incorporates soft information (e.g., human factors, personal experiences, personal opinions, hunches, and personal insights) and do not rely on rigorous mathematical computations. There are many types of qualitative forecasting methods, some informal and some structured. Regardless of how structured the process is, however, remember that these models are based on subjective opinion and are not mathematical in nature [4]. Qualitative forecasting methods include executive opinion, market research, and Delphi method. Quantitative methods involve either the projection of historical data or the development of associative models that attempt to utilize causal (explanatory) variables to make a forecast [5]. Quantitative techniques mainly analyze objective, or hard, data and usually avoid personal biases that sometimes contaminate qualitative methods.

Quantitative methods can also be divided into two categories: time series models and causal models. Although both are mathematical, the two categories differ in their assumptions and in the manner in which a forecast is generated [6].

Time series models assume that all the information needed to generate a forecast is contained in the time series of data. A time series is a series of observations taken at regular intervals over a specified period of time. Time series analysis assumes that we can generate a forecast based on patterns in the data. As a forecaster, you would look for patterns such as trend, seasonality, and cycle and use that information to generate a forecast.
Causal models, sometimes called associative models, use a very different logic to generate a forecast. They assume that the variable we wish to forecast is somehow related to other variables

in the environment. The forecaster's job is to discover how these variables are related in mathematical terms and use that information to forecast the future.
Time series models include the following:

(a) Naïve Methods. A naive forecast uses a single previous value of a time series as the basis of a forecast. The naive approach can be used with a stable series (variations around an average), with seasonal variations, or with trend. For data with trend, the forecast is equal to the last value of the series plus or minus the difference between the last two values of the series. For example, suppose the last two values were 50 and 53. The next forecast would be 56 [7]:

Table 2.1. Showing naïve method

Period	Actual	Change from previous value	Forecast
1	50		
2	53	+3	
3			53+3 = 56

(b) Simple Moving Average. One weakness of the naive method is that the forecast just traces the actual data, with a lag of one period; it does not smooth at all. But by expanding the amount of historical data a forecast is based on, this difficulty can be overcome. A simple moving average forecast uses a number of the most recent actual data values in generating a forecast. The simple moving average forecast can be computed using the following equation:

$$F_t = SMAn = \frac{\sum_{i=1}^{n} A_{t-i}}{n} = \frac{A_{t-n} + \ldots + A_{t-i} + A_{t-1}}{n}$$

Where,

F_t = Forecast for time period t,
SMAn = n period simple moving average
A_{t-i} = Actual value in period t-i
n = Number of periods (data points) in the simple moving average

(c) Weighted Moving Average. A weighted average is similar to a simple moving average, except that it assigns more weight to the most recent values in a time series.

$$F_t = w_t(A_t) + w_{t-1}(A_{t-1}) + \ldots + w_{t-n}(A_{t-n})$$

Where,

W_t = Weight for the period t
W_{t-1} = Weight for period t-1, etc
A_t = Actual value in period t
A_{t-1} = Actual value for period t-1, etc

(d) Exponential Smoothing. Exponential smoothing is a sophisticated weighted averaging method that is still relatively easy to use and understand. Each new forecast is based on the previous forecast plus a percentage of the difference between that forecast and the actual value of the series at that point. That is:

Next forecast = Previous forecast + α(Actual – Previous forecast)

Where (Actual – Previous forecast) represents the forecast error and α is a forecasting error. More concisely,

$$F_t = F_{t-1} + \alpha(A_{t-1} - F_{t-1})$$

Where,

F_t = Forecast for period t
F_{t-1} = Forecast for the previous period
α = Smoothing constant (percentage)
A_{t-1} = Actual demand for the previous period

Exponential smoothing is one of the most widely used techniques in forecasting, partly because of its ease of calculation and partly because of the ease with which the weighting scheme can be altered - simply by changing the value of α [7].

(e) Trend Analysis. Analysis of trend involves developing an equation that will suitably describe trend. The trend may be linear or nonlinear. One way to describe the trend component is to fit a line visually to a set of points on a graph. Any given graph, however, There are two important techniques that can be used to develop forecasts when trend is present. One involves use of a trend equation; the other is an extension of exponential smoothing.

❖ Trend Line Equation. A linear equation has the form

$$F_t = a + b_t$$

Where,

F_t = Forecast for period t
a = Value of F_t at t = 0
b = Slope of the line
t = Specified number of time periods from t = 0

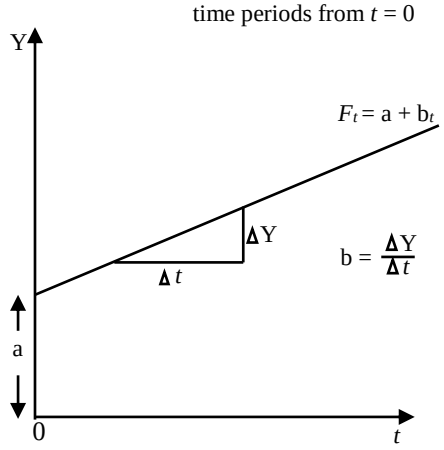


Figure 2.2. Showing trend line equation

- Trend-Adjusted Exponential Smoothing. A variation of simple exponential smoothing can be used when a time series exhibits a linear trend. It is called trend-adjusted exponential smoothing or, sometimes, double smoothing, to differentiate it from simple exponential smoothing, which is appropriate only when data vary around an average or have step or gradual changes. If a series exhibits trend, and simple smoothing is used on it, the forecasts will all lag the trend: If the data are increasing, each forecast will be too low; if decreasing, each forecast will be too high. The trend-adjusted forecast (TAF) is composed of two elements: a smoothed error and a trend factor [7].

$$TAF_{t+1} = S_t + T_t$$

Where,
S_t = Previous forecast plus smooth error
T_t = Current trend estimate
and,
$S_t = TAF_t = \alpha(A_t - TAF_t)$
$T_t = T_{t-1} + \beta(TAF_t - TAF_{t-1} - T_{t-1})$
Where,
α = Smoothing constant for averaging
β = Smoothing constant for trend

In order to use this method, one must select values of α and β (usually through trial and error) and make a starting forecast and an estimate of trend. Unlike a linear trend line, trend-adjusted smoothing has the ability to adjust to changes in trend. Of course, trend projections are much simpler with a trend line than with trend-adjusted forecasts, so a manager must decide which benefits are most important when choosing between these two techniques for trend.

Causal (Associative) Models include the following:

(a) Linear Regression Models. Linear regression refers to the special class of regression where the relationship between variables forms a straight line. The linear regression line is of the form Y = a + bX, where Y is the value of the dependent variable that we are solving for, a is the Y intercept, b is the slope, and X is the independent variable. (In time series analysis, X is units of time). Linear regression is useful for long-term forecasting of major occurrences and aggregate planning. The major restriction in using linear regression forecasting is, as the name implies, that past data and future projections are assumed to fall about a straight line.

(b) Multiple Regression Models. Multiple regression develops a relationship between a dependent variable and multiple independent variables. The general formula for multiple regression is as follows:

$$Y = B_0 + B_1X_1 + B_2X_2 + \ldots + B_kX_k$$

Where,
Y = Dependent variable
B_0 = The Y intercept
$B_1 \ldots B_k$ = Coefficients that represents the influence of the independent variables on the dependent variable.
$X_1 \ldots X_k$ = Independent variables.

Multiple regression is a powerful tool for forecasting and should be used when multiple factors influence the variable that is being forecast.

2.3 Causes of Inaccuracy in Forecasting

Accuracy and control of forecasts is a vital aspect of forecasting, so forecasters want to minimize forecast errors. However, the complex nature of most real-world variables makes it almost impossible to correctly predict future values of those variables on a regular basis. Moreover, because random variation is always present, there will always be some residual error, even if all other factors have been accounted for. Consequently, it is important to include an indication of the extent to which the forecast might deviate from the value of the variable that actually occurs. This will provide the forecast user with a better perspective on how far off a forecast might be [7]. Forecast error is the difference between the value that occurs and the value that was predicted for a given time period. Error is the difference between the value that occurs and the value that was predicted for a given time period. Hence,

Error = Actual - Forecast:
$E_t = A_t - F_t$
where
t = Any given time period.

Error can occur in two ways:

(a) BIAS: A bias is a constant deviation from the mean in one direction (high or low). In terms of forecasting bias is the tendency of the forecast to be either above or below the actual observations. With this concept, if the computed bias – ve, the forecast is consistently too low; if the computed bias is +ve, the forecast is consistently too high. Bias is calculated as the total error divided by the number of periods.

$$\text{Bias} = \frac{\sum(\text{Actual} - \text{Forecast})}{\text{No. of Periods}}$$

(b) RANDOM VARIATION: In a given period, actual demand will vary about the average demand. The differences are random variations. The variability will depend on the demand pattern of the product. Some products will have a stable demand, and the variation will not be large. Others will be unstable and will have a large variation.

2.4 Improving Forecast Accuracy

2.4.1Tracking Error Signal

When there is a difference between forecast and actual values, one problem is to identify whether the difference is caused by random

variation or is due to a bias in the forecast. Forecast bias is a persistent tendency for a forecast to be over or under the actual value of the data. We cannot do anything about random variation, but bias can be corrected. One way to control forecast bias is to use a tracking signal. A tracking signal is a tool used to monitor the quality of the forecast. It is computed as the ratio of the algebraic sum of the forecast errors divided by MAD:

$$\text{Tracking Signal} = \frac{\text{Algebraic sum of forecast errors}}{\text{MAD}}$$

Or

$$\text{Tracking Signal} = \frac{\sum(\text{Actual}_t - \text{Forecast}_t)}{\text{MAD}}$$

Where MAD is the mean absolute deviation given mathematically as:

$$\frac{\sum|\text{Actual}_t - \text{Forecast}_t|}{n}$$

As the forecast errors are summed over time, they can indicate whether there is a bias in the forecast. To monitor forecast accuracy, the values of the tracking signal are compared against predetermined limits. These limits are usually based on judgment and experience and can range from ±3 to ±8.

2.5 Role of Computer and Software in Forecasting

Computers allow you to take all the vast amounts of data about the previous year(s), quarter(s) or month(s) and compile them into meaningful information that you can then use to reveal important trends in your organization. Compiling all of the data is a daunting task that can only be handled accurately by intuitive forecasting software. Such software make the processes much simpler and more manageable. Through this synthesis of data, you can find out what is and what is not working in your organization. You can also easily track and pinpoint exactly where things are beginning to go wrong. Data is only useful in as much as you can interpret and use it. To interpret and use data, the right forecasting tools and the right methods are required.

Computers play an important role in preparing forecasts based on quantitative data. Their use allows managers to develop and revise forecasts quickly, and without the burden of manual computations.

Today much commercial forecasting is performed using computer software. Many software packages can be used for forecasting. Some can handle thousands of variables and manipulate huge databases. Others specialize in one forecasting model. Consequently, it may be difficult to select the right forecasting software. Most forecasting software packages fall into one of three categories: (1) spreadsheets, (2) statistics packages, and (3) specialty forecasting packages.

3. METHODOLOGY

The study utilized survey research design to obtain relevant data. Three hypotheses were postulated to guide us in the work. The method of research involved administering of questionnaires and subsequent analysis of the results of the questionnaires using chi-squared goodness of fit tests. The results obtained from the chi-squared analysis were used to test the hypotheses.

3.1 Hypotheses

To guide our work in this study, the following hypotheses were posited.

i. There is significant relationship between use of forecasting software and minimizing the impact of forecast error on government's monetary and fiscal policies.

ii. Use of forecasting software makes computation of seasonal variations and large historical data less cumbersome.

iii. Quantitative methods of forecasting produce less error than qualitative methods of forecasting.

3.2 Source of Data

We studied the effect of software on minimizing the impact of forecast error on governments' monetary and fiscal policies with data from two main sources:

Primary sources: Questionnaires are used to obtain relevant data from stakeholders in the forecasting industry. A total of 128 questionnaires covering 10 questions were delivered by hand to the stakeholders in forecasting industry which included economic forecasters, software developers, lecturers, and university students. Out of this number, 120 questionnaires were completed and returned. The questions sought, among others, the views of the respondents on the relationship between forecasting software and minimizing the impact of forecast error on government's monetary and fiscal policies.

Secondary Source: Relevant information were drawn from articles and books written by other professionals in the IT and forecasting industry.

3.3 Data Analysis and Result Presentation

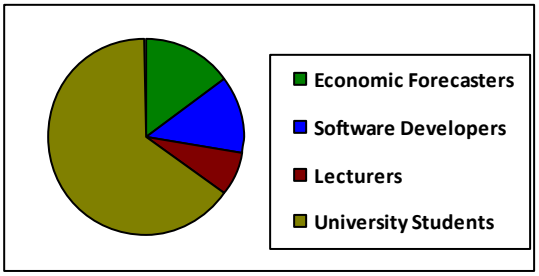
Quantitative data obtained from primary source were analyzed using the chi-squared test. Table 3.1 and figure 3.1 both show the occupational distribution of respondents.

Table 3.1. Occupational distribution of respondents

S/n	Respondents' Occupation	Number	Percentage
1.	Economic forecasters	18	15%
2.	Software Developers	15	12.5%
3.	Lecturers	9	7.5%
4.	University Students	78	65%
	Total:	**120**	**100%**

Source: Field study (2016)

Figure 3.1. Pie chart showing occupational distribution of respondents



Source: Field study (2016)

Table 3.2. Questions, responses and X2 values from the Chi-squared analysis

S/n	Questions	O_i	E_i	$(O_i - E_i)$	$(O_i - E_i)^2$	$(O_i - E_i)^2/E_i$	$X^2 = \sum[(O_i - E_i)2/E_i]$
1.	Challenges in economic forecasting, such as increasing accuracy and reducing bias, are best met through effective management of the forecasting process. • SA • A • U • D • SD	80 22 0 12 6	24 24 24 24 24	56 -2 -24 -12 -18	3136 4 576 144 324	130.67 0.17 24 6 13.5	174.34
2.	Effective management of economic forecasting process requires an understanding of the realities, limitations, and principles fundamental to the process. • SA • A • U • D • SD	78 22 2 12 6	24 24 24 24 24	54 -2 -22 -12 -18	2916 4 484 144 324	121.5 0.17 20.17 6 13.5	161.34
3.	When management lacks a grasp of basic concepts like randomness, variation, uncertainty, and forecastability, the organization is apt to squander time and resources on expensive and unsuccessful fixes. • SA • A • U • D • SD	92 18 0 6 4	24 24 24 24 24	68 -6 -24 -18 -20	4624 36 576 324 400	192.67 1.5 24 13.5 16.67	248.34
4.	Forecast accuracy is ultimately limited by the nature of what we trying to forecast. • SA • A • U • D • SD	74 24 4 10 8	24 24 24 24 24	50 0 -20 -14 -16	2500 0 400 196 256	104.17 0 16.67 8.17 10.67	139.68
5.	Accurate forecasts improve government's chances of making the right decisions about its monetary and fiscal policies. • SA • A • U • D • SD	96 18 0 4 2	24 24 24 24 24	72 -6 -24 -20 -22	5184 36 576 400 484	216 1.5 24 16.67 20.17	278.34

6.	The "perfect forecast" is the one that has enough information to improve manage-ment's decisions under conditions of uncertainty.						
	• SA	86	24	62	3844	160.17	
	• A	22	24	-2	4	0.17	
	• U	2	24	-22	484	20.17	
	• D	7	24	-17	289	12.04	
	• SD	3	24	-21	441	18.38	210.93
7.	In order to be useful to manage-ment decision making, forecasts should be timely, reliable and accurate.						
	• SA	94	24	70	4900	204.17	
	• A	22	24	-2	4	0.17	
	• U	1	24	-23	529	22.04	
	• D	2	24	-22	484	20.17	
	• SD	1	24	-23	529	22.04	268.59
8.	Quantita-tive methods of forecasting produce less error than qualitative methods because of elimination of human bias.						
	• SA	102	24	78	6084	253.5	
	• A	4	24	-20	400	16.67	
	• U	2	24	-22	484	20.17	
	• D	8	24	-16	256	10.67	
	• SD	4	24	-20	400	16.67	317.68
9.	Manual computa-tions of seasonal variations and large historical data are a bit cumber-some, so the use of forecasting software is preferable.						
	• SA	118	24	94	8836	368.17	
	• A	2	24	-22	484	20.17	
	• U	0	24	-24	576	24	
	• D	0	24	-24	576	24	
	• SD	0	24	-24	576	24	460.34
10.	Effective manage-ment of economic forecasting process is best achieved through the use of computer software.						
	• SA	98	24	74	5476	228.17	
	• A	14	24	-10	100	4.17	
	• U	2	24	-22	484	20.17	
	• D	4	24	-20	400	16.67	
	• SD	2	24	-22	484	20.17	289.35

3.4 Test of Hypotheses

3.4.1 Hypothesis One

H_0: There is no significant relationship between use of forecasting software and minimizing the impact of forecast error on government's monetary and fiscal policies.

H_1: There is significant relationship between use of forecasting software and minimizing the impact of forecast error on government's monetary and fiscal policies.

Relevant in testing hypothesis one is question 10 of the questionnaire.

From the chi-squared analysis on table 3.2, $X^2 = \sum[(O_i-E_i)^2/E_i]$ for question 10 is 289.35.
Our degree of freedom (d. f.) = (n - 1) = (5 – 1) = 4 and our level of significance is 0.05

Decision

Tabulated value of X^2 (X^2_{Tab}) at 4 d. f. and 0.05 = 9.488
The calculated value of X_2 (X^2_{Cal}) is 289.35
$X^2_{Cal} > X^2_{Tab}$
The decision rule states that if X^2_{Cal} is greater than X^2_{Tab}, we should reject the null hypothesis (H_0) and accept the alternative hypothesis (H_1).

We therefore reject the null hypothesis (H_0) and accept the alternative hypothesis (H_1).

3.4.2 Hypothesis Two

H_0: Use of forecasting software does not make computation of seasonal variations and large historical data less cumbersome.

H_1: Use of forecasting software makes computation of seasonal variations and large historical data less cumbersome.

Relevant in testing hypothesis two is question 9 of the questionnaire.

From the chi-squared analysis on table 3.2, $X^2=\sum[(O_i-E_i)^2/E_i]$ for question 9 is 460.34.
Our degree of freedom (d. f.) = (n - 1) = (5 – 1) = 4 and our level of significance is 0.05

Decision
Tabulated value of X^2 (X^2_{Tab}) at 4 d. f. and 0.05 = 9.488
The calculated value of X_2 (X^2_{Cal}) is 460.34
$X^2_{Ca}l > X^2_{Tab}$
The decision rule states that if X_{2Cal} is greater than X^2_{Tab}, we should reject the null hypothesis (H_0) and accept the alternative hypothesis (H_1).

We therefore reject the null hypothesis (H_0) and accept the alternative hypothesis (H_1).

3.4.3 Hypothesis Three

H_0: Quantitative methods of forecasting do not produce less error than qualitative methods of forecasting.

H_1: Quantitative methods of forecasting produce less error than qualitative methods of forecasting.

Relevant in testing hypothesis three is question 8 of the questionnaire.

From the chi-squared analysis on table 3.2, $X^2=\sum[(O_i-E_i)^2/E_i]$ for question 8 is 317.68.
Our degree of freedom (d. f.) = (n - 1) = (5 – 1) = 4 and our level of significance is 0.05
Decision
Tabulated value of X^2 (X^2_{Tab}) at 4 d. f. and 0.05 = 9.488
The calculated value of X_2 (X^2_{Cal}) is 317.68
$X^2_{Cal} > X^2_{Tab}$
The decision rule states that if X^2_{Cal} is greater than X^2_{Tab}, we should reject the null hypothesis (H_0) and accept the alternative hypothesis (H_1).

We therefore reject the null hypothesis (H_0) and accept the alternative hypothesis (H_1).

4. FINDINGS/DISCUSSION

4.1 Findings

The findings from this study, among others, reveal that:

i. Challenges in economic forecasting, such as increasing accuracy and reducing bias, are best met through effective management of the forecasting process.

ii. Effective management of economic forecasting process is best achieved through the use of computer software.

iii. Manual computations of seasonal variations and large historical data are a bit cumbersome, so the use of forecasting software is preferable.

4.2 Discussion

In order to get necessary data for this study, the survey method was used. Data obtained from the respondents were analyzed using the chi-square goodness of fit tests. To guide us in the study, three hypotheses were raised: there is significant relationship between the use of forecasting software and minimizing the impact of forecast error on government's monetary and fiscal policies; use of forecasting software makes computation of seasonal variations and large historical data less cumbersome; and quantitative methods of forecasting produce less error than qualitative methods of forecasting. The tests of the hypotheses showed that: there is significant relationship between the use of forecasting software and minimizing the impact of forecast error on government's monetary and fiscal policies; the use of forecasting software makes computation of seasonal variations and large historical data less cumbersome; and that quantitative methods of forecasting produce less error that qualitative methods.

All the questions listed in table 4 for the opinion of the respondents had calculated chi-square values greater than their corresponding table values for chi-squares for the data sets we are analyzing at 4 degree of freedom and p = 0.05 level of significance. These show that there are significant differences between the data sets that cannot be due to chances alone.

5. CONCLUSION

Forecasts are not perfect; actual results usually differ from predicted values; the presence of randomness and bias precludes a perfect forecast. As a result of this, forecasters always look for innovative ways of minimizing forecast errors. The goal of an effective forecasting system is to generate good forecasts on the average over time and to keep forecast errors as low as possible. To this end, this study was undertaken to:

i. identify the various techniques used in economic forecasting;
ii. to indentify the causes of error in economic forecasting;
iii. to identify ways of improving forecast accuracy and minimizing forecast error; and
iv. to focus on ways of minimizing the impact of forecast error on government's monetary and fiscal policies.

In conclusion therefore, producing economic forecast is necessary, as they are essential underlying features of governments' budgetary, monetary, and other fiscal policies. In doing this, policy makers should focus on ways of minimizing the impact of forecast errors on such fiscal and monetary policies. As shown in this study, an effective way of minimizing the impact of forecast error on government's monetary and fiscal policy lies in the use of effective software driven forecasting system.

6. REFERENCES

[1] Diebold, F. X. 2015. Forecasting in economics, business, finance and beyond. University of Pennsylvania, USA.

[2] Gilliland, M., Tashman, L., and Sglavo, U. 2015. Business forecasting: practical problems and solutions. John Wiley & Sons, New Jersey, USA.

[3] Hyndman, R. J. 2014. Forecasting principles and practice. University of Western Australia, Australia.

[4] Reid, R. D. and Sanders, N. R. 2011. Operations management. An integrated approach. [Fourth Edition]. John Wiley & Sons Inc., New Jersey, USA

[5] Stevenson, W. J. 2012. Operations management [eleventh edition]. McGraw-Hill, New York City, USA.

[6] Reid, R. D. and Sanders, N. R. 2011. Operations Management. An integrated Approach. [Fourth Edition]. John Wiley & Sons Inc., New Jersey, US

[7] Stevenson, W. J. 2012. Operations Management [eleventh edition]. McGraw-Hill, New York City, USA.

International Journal of Computer Applications Technology and Research
Volume 6–Issue 3, 168-171, 2017, ISSN:-2319–8656

Clustering Students of Computer in Terms of Level of Programming

Shirin Esmailbegi Kermani
Department of Computer,Kerman Branch,Islamic Azad University, Kerman,Iran

Abbas Rezaie
Department of Computer,Kerman Branch,Islamic Azad University, Kerman,Iran

Abstract: Educational data mining (EDM) is one of the applications of data mining. In educational data mining, there are two key domains, i.e. student domain and faculty domain. Different type of research work has been done in both domains.
In existing system the faculty performance has calculated on the basis of two parameters i.e. Student feedback and the result of student in that subject. In existing system we define two approaches one is multiple classifier approach and the other is a single classifier approach and comparing them, for relative evaluation of faculty performance using data mining
Techniques. In multiple classifier approach K-nearest neighbor (KNN) is used in first step and Rule based classification is used in the second step of classification while in single classifier approach only KNN is used in both steps of classification.
But in proposed system, I will analyse the faculty performance using 4 parameters i.e., student complaint about faculty, Student review feedback for faculty, students feedback, and students result etc.
For this proposed system I will be going to use opinion mining technique for analyzing performance of faculty and calculating score of each faculty.

Keywords: clustering algorithm, Data mining, Educational data mining, classification

1. INTRODUCTION

1. Data Mining is the process of analyzing data from different perspectives and summarizing the result as useful information. Educational Data mining is an emerging research field in data mining and continuously growing with a boom. Incalculable data mining techniques have been applied to the variety of educational data [15].
2. For the better decision making in the learning environment, EDM consists of four phases: Data Preprocessing Data Validation Data Prediction Decision making process.
3. This study investigates the education domain for relative evaluation of faculty performance on the basis of two parameters student feedback and result in a course taught by faculty. Classification technique of data mining is applied to fulfill the objective. The goal of classification technique is to place an object into a category based on the characteristics of the object.
4. The whole paper is organized as: In section 1 Introduction about DM, EDM and some idea of relative problem is mentioned. In section 2 overview of related work in which educational data mining on student domain and faculty domain is discussed from previous research. In section 3 proposed brings up where educational data mining is presented in faculty domain. The proposed framework is explained in section 4.

For the better decision making in the learning environment, EDM consists of four phases:
Data Pre-processing
Data Validation
Data Prediction
Decision making process

2. LITERATURE REVIEW

2.2 A Comprehensive Study of Educational Data Mining:
Author Jasvinder Kumar suggests that, educational data mining is a new discipline in research community that applies various tools and techniques of data mining (DM) to explore data in the field of education [2]. This discipline helps to learn and develop models for the growth of education environment. It provides decision makers a better understanding of student learning and the environment setting in as of EDM. It also highlights the opportunities for future research. Educational data Mining (EDM) has been evolved as multidisciplinary scientific learning area, rich in data, methods, tools and techniques used to provide better learning environment for educational users in educational context. This paper integrates all the modules of EDM required to facilitate the objectives of educational research. Lastly it shows that, there are many more research topics that exist in this domain.[2] Utilization of data mining techniques within education environment requires a joint effort by the ICT specialists, educationists and the learners.

2.3 Educational Data Mining: a Case Study-
Author Kalina suggest in this paper, Author show how using data mining algorithms can help discovering pedagogically relevant knowledge contained in databases obtained from Web-based educational systems. These findings can be used both to help teachers with managing their class, understand their students learning and reflect on their teaching and to support learner reflection and provide proactive feedback to learners. In this paper, Author has shown how the discovery of different patterns through different data mining algorithms and visualization techniques suggests to us a simple pedagogical policy. Data exploration focused on the number of attempted exercises combined with classification led us to identify

students at risk, those who have not trained enough. Clustering and cluster visualization led us to identify a particular behavior among failing students, when students try out the logic rules of the pop-up menu of the tool. As in [5], a timely and appropriate warning to students at risk could help preventing failing in the final exam. Therefore it seems to us that data mining has a lot of potential for education, and can bring a lot of benefits in the form of sensible, easy to implement pedagogical policies as above. The way Author have performed clustering may seem rough, as only few variables, namely the number and type of mistakes, the number of exercises have been used to cluster students in homogeneous groups. This is due to our particular data. All exercises are about formal proofs. Even if they differ in their difficulty, they do not fundamentally differ in the concepts students have to grasp. Author have discovered a behavior rather than particular abilities. In a different context, clustering students to find homogeneous groups regarding skills should take into account answers to a particular set of exercises. Currently, Author are doing research work along these lines.

2.4 Extraction of rules based on students questionnaires:

Author Manolis Chalaris; suggest there are many students in the Greek Higher Education that are still "lingering" in their Departments beyond the six years. The length of studies beyond 6 years has not been justified, and this study focuses on this problem. Author also studies another problem: The percentage of graduates scoring about 8.5/10 or more is extremely low [8]. Association rules mining is a well known data analysis method for extracting associations between data in a wide range of different fields. In this paper, Author focus on the generation of the appropriate association rules based on students' questionnaires in Higher Education. A sample of 50.000 questionnaires was filled by 10.000 students in the TEI of Athens. Various interesting rules could be extracted related to learning goals, practices, years required for graduation, etc. These rules and clustering techniques could be used for solving the problem of the students that are still "lingering", and the problem of the low "scoring" of the graduates. In this section, Author use the simple example of the section 3 in order to present the potential impact of these results in Higher Education, and how the described technique could be competitive.

2.5 Predicting Students academic Performance using Artificial Neural Network: A Case Study of an Engineering Course:

Author V.O. Oladokun suggest that , The observed poor quality of graduates of some Nigerian Universities in recent times has been partly traced to inadequacies of the National University Admission Examination System. In this study an Artificial Neural Network (ANN) model, for predicting the likely performance of a candidate being considered for admission into the university was developed and tested. Various factors that may likely influence the performance of a student were identified. Such factors as ordinary level subjects scores and subjects' combination, matriculation examination scores, age on admission, parental background, types and location of secondary school attended and gender, among others, were then used as input variables for the ANN model. A model based on the Multilayer Preceptor Topology was developed and trained using data spanning five generations of graduates from an Engineering Department of University of Ibadan, Nigeria's first University. Test data evaluation shows that the ANN model is able to correctly predict the performance of more than 70% of prospective students. This study has shown the potential of the artificial neural network for enhancing the effectiveness of a university admission system. The model was developed based on some selected input variables from the pre admission data of five different sets of university graduates. It achieved an accuracy of over 74%, which shows the potential efficacy of Artificial Neural Network as a prediction tool and a selection criterion for candidates seeking admission into a university. One limitation of this model stems from the fact that not all the relevant performance influencing factors are obtainable from the pre-admission record forms filled by the students. A model incorporating the use of results from a carefully designed oral interview administered to the students may likely be an improvement over the present model. Also the extension this research to non-engineering departments is recommended. The current admissions system should be reviewed in order to improve the standard of candidates being admitted into the institution. A more adequate ANN may be very useful for such an exercise.[11]

2.6 Survey on Decision Tree Classification algorithms for the Evaluation of Student Performance:

Author Anjou Rathee, suggest that Now days, the amount of data stored in educational database is increasing rapidly. These databases contain hidden information for improvement of student's performance. Classification of data objects is a data mining and knowledge management technique used in grouping similar data objects together. There are many classification algorithms available in literature but decision tree is the most commonly used because of its ease of implementation and easier to understand compared to other classification algorithms. The ID3, C4.5 and CART decision tree algorithms has been applied on the data of students to predict their performance. In this paper, all the algorithms are explained one by one. Performance and results are compared of all algorithms and evaluation is done by already existing datasets. All the algorithms have a satisfactory performance but accuracy is more witnessed in case of C4.5 algorithm. In this paper, three existing decision tree algorithms have been applied on the educational data for predicting the student's performance in examination. All the algorithms are applied on student's internal assessment data to predict their performance in the final exam. The efficiency of various decision tree algorithms can be analyzed based on their accuracy and time taken to derive the tree. The predictions obtained from the system have helped the tutor to identify the weak students and improve their performance. [14] C4.5 is the best algorithm among all the three because it provides better accuracy and efficiency than the other algorithms. Still effective algorithms for decision tree should be developed.

2.7 Survey on Decision Tree Classification algorithms for the Evaluation of Student Performance:

Author Anjou Rathee, suggest that Now days, the amount of data stored in educational database is increasing rapidly. These databases contain hidden information for improvement of student's performance. Classification of data objects is a data mining and knowledge management technique used in grouping similar data objects together. There are many classification algorithms available in literature but decision tree is the most commonly used because of its ease of implementation and easier to understand compared to other classification algorithms. The ID3, C4.5 and CART decision tree algorithms has been applied on the data of students to predict their performance. In this paper, all the algorithms are explained one by one. Performance and results are compared of all algorithms and evaluation is done by already existing datasets. All the algorithms has a satisfactory performance but accuracy is more witnessed in case of C4.5 algorithm. In this paper, three existing decision tree algorithms have been applied on the educational data for predicting the student's performance in examination. All the algorithms are applied on student's internal assessment data to predict their performance in the final exam. The efficiency of various decision tree algorithms can be analyzed based on their accuracy and time taken to derive the tree. The predictions obtained from the system have helped the tutor to identify the weak students and improve their performance.[14] C4.5 is the best algorithm among all the three because it provides better accuracy and efficiency than the other algorithms. Still effective algorithms for decision tree should be developed.

a. Rule Mining Framework for Students Performance Evaluation:

Author P. Ajith, B. Tejaswi, M.S.S.Sai suggest that Academic Data Mining used many techniques such as Decision Trees, Neural Networks, Naïve Bayes, K- Nearest neighbor, and many others. Using these techniques many kinds of knowledge can be discovered such as association rules, classifications and clustering. The discovered knowledge can be used for prediction and analysis purposes of student patterns. Prior approaches used decision tree classifications optimized with ID3 algorithms to obtain such patterns. Among sets of items in transaction databases, Association Rules aims at discovering implicative tendencies that can be valuable information for the decision-maker which is absent in tree based classifications. So we propose a new interactive approach to prune and filter discovered rules. First, Author propose to integrate user knowledge in the post processing task. Second, Author propose a Rule Schema formalism extending the specifications to obtain association rules from knowledge base. Furthermore, an interactive framework is designed to assist the user throughout the analyzing task. Applying our new approach to discover the likelihood of students deviations / requiring special attention is organized and efficient providing more insight by considering more information.[12]

3. PROPOSED WORK:

Similar to this, the Data Mining technique can also help the education industry to achieve better goals. Educational institutes can apply data mining techniques for analyzing student performance, faculty performances and the overall performance of institutes. Institutes can study comparative performances with other institutes to compete with and lead other institutes and improve their ratings.

Data Collection
Data Preprocessing
Refined Data
KNN Algorithm
4 Parameters used(Calculate result for each parameter)
Sum of all parameters
Rule based classification algorithm
Analysis of all subject and class of each faculty result

Figure 1 Flow of System

So, By using multiple classifier approach and cure clustering algorithm the performance of the system is better than the existing one. Here KNN is used for mapping purpose. And rule based algorithm is used for deciding the rules on the data. Following figure shows the flow of the proposed system that will be implemented.

4. CONCLUSIONS

The overall performance of multiple classifier approach is better than the single classifier approach. In the second step of multiple classifier approach we Rule-based classification have been used where the authors define their own rules for classification, which make the difference from single classifier approach. In single classifier approach we sum up the both parameters scores which may restrict the performance of this approach. Future work can enhance the performance of the approach by considering some more parameters according to the Requirement of the organization. Secondly, in the present work, author has taken only Computer Science and engineering department faculties. Future work can be enhanced for all branches of the college for overall performance enhancement of the college.

5. REFERENCES

[1] N.Aggrawal et.al,"Analysis the effect of data mmmg techniques on Database" Elsevier, Advances in Engineering Software 47 (2012) 164-169.

[2] N.Ahmad and S.Shamsuddin," A Comparative Analysis of Mining Techniques for Automatic Detection of Students Learning Styles," IEEE, 978-1-4244-8136-1
(2010).

[3] Z.Abdullah et.al," Extracting highly positive association rules from students enrollment data", Elsevier, Procedia - Social and Behavioral Science 28 (2011) 107-111.

[4] K.Bunkar et.al," Data Mining: Prediction for

Performance Improvement of Graduation Student using Classification", IEEE, 978-4673-1989-8 (2012).
[5] B.Bidgoli et.al,"Predicting Student Performance: An Application of Data Mining Methods with an Educational Web-Based System", 33rd ASEE/IEEE Frontiers in Education Conference T2A-13, November 5-8, 2003.
[6] N.Delavari et.al,"A New Model for Using Data Mining Technology in Higher Education Systems", IEEE, 0-7803-8596-9 (2004).
[7] E.Garcia and P.Mora,"Model Prediction of academic performance for fust year students", IEEE, 978-0-7695- 4605-6 (2011).
[8] S.Liao et.al,"Data Mining techniques and applicationsA decade review from 2000-2011",Elsevier, Expert System with Application 39 (2012) 11303-11311.
[9] LiPing and DuanFu,"The Disposal of Incomplete Classification Data in Teaching Evaulation System", IEEE, 2009 Third International Symposium on Intelligent Information Technology Application, 978-7695-3859-4
[10] X. Long and y.wu,"Application of decision tree in student achievement evaluation", IEEE, 2012 International Conference on Computer Science and Electronics Engineering, 978-7695-4647-6.
[11] S. Mardikyan and B. Badur,"Analysis Teaching Performance of Instructors Using Data Mining Techniques" ,Information in Education, 2011, Vol. 1 0, No.2,245-257.
[12] P.Qingxian et.al,"data millllg and application of teaching evaluation based on association rules", IEEE, Proceeding of 2009 4th International Conference on Computer Science and Education, 978-4244-3521-0
[13] Al.Radaishel et.al,"Mining student data using decision tree", in the proceeding of the International Arab conference on information technology, Jordan, 2006.
[14] C.Romero and S.Ventura", Educational data mining: A survey from 1995 to 2005". Elsevier, Expert System with Application 33 (2007) 135-146.
[15] C.Singh et.al,"Extraction and Analysis of Faculty Performance of Management Discipline from Student Feedback Using Clustering and Association rule mining techniques", IEEE, 978-4244-8679-3 (2011).
[16] K. Shaban and M. Abdulwahed,"Research-Based Learning in Computing Courses for Senior Engineering Students", IEEE International Conference on Teaching, assessment and Learning for Engineering (TALE) 2012, 978-1-4673-2418-2.
[17] H.Wang and R.Zhong,"Application of Association Rules in College Teaching",IEEE, 2011 International Conference on Internet Computing and Information Services, 978-0-7695-4539-4.