David C. Wyld
Natarajan Meghanathan (Eds)

Computer Science & Information Technology

Fifth International Conference on Advances in Computing and Information Technology (ACITY 2015)
Chennai, India, July 25~26, 2015

AIRCC

Volume Editors

David C. Wyld,
Southeastern Louisiana University, USA
E-mail: David.Wyld@selu.edu

Natarajan Meghanathan,
Jackson State University, USA
E-mail: nmeghanathan@jsums.edu

ISSN: 2231 - 5403
ISBN: 978-1-921987-41-0
DOI : 10.5121/csit.2015.51301 - 10.5121/csit.2015.51318

Typesetting: Camera-ready by author, data conversion by NnN Net Solutions Private Ltd., Chennai, India

Preface

The Fifth International Conference on Advances in Computing and Information Technology (ACITY 2015) was held in Chennai, India, during July 25~26, 2015. The Fifth International Conference on Digital Image Processing and Pattern Recognition (DPPR 2015), The Sixth International Conference on VLSI (VLSI 2015), The Second International Conference on Wireless and Mobile Network (WiMNET 2015), The Fifth International Conference on Artificial Intelligence, Soft Computing and Application (AIAA 2015) and The Second International Conference on Computer Networks & Data Communications (CNDC 2015) were collocated with the ACITY-2015. The conferences attracted many local and international delegates, presenting a balanced mixture of intellect from the East and from the West.

The goal of this conference series is to bring together researchers and practitioners from academia and industry to focus on understanding computer science and information technology and to establish new collaborations in these areas. Authors are invited to contribute to the conference by submitting articles that illustrate research results, projects, survey work and industrial experiences describing significant advances in all areas of computer science and information technology.

The ACITY-2015, DPPR-2015, VLSI-2015, WiMNET-2015, AIAA-2015, CNDC-2015 Committees rigorously invited submissions for many months from researchers, scientists, engineers, students and practitioners related to the relevant themes and tracks of the workshop. This effort guaranteed submissions from an unparalleled number of internationally recognized top-level researchers. All the submissions underwent a strenuous peer review process which comprised expert reviewers. These reviewers were selected from a talented pool of Technical Committee members and external reviewers on the basis of their expertise. The papers were then reviewed based on their contributions, technical content, originality and clarity. The entire process, which includes the submission, review and acceptance processes, was done electronically. All these efforts undertaken by the Organizing and Technical Committees led to an exciting, rich and a high quality technical conference program, which featured high-impact presentations for all attendees to enjoy, appreciate and expand their expertise in the latest developments in computer network and communications research.

In closing, ACITY-2015, DPPR-2015, VLSI-2015, WiMNET-2015, AIAA-2015, CNDC-2015 brought together researchers, scientists, engineers, students and practitioners to exchange and share their experiences, new ideas and research results in all aspects of the main workshop themes and tracks, and to discuss the practical challenges encountered and the solutions adopted. The book is organized as a collection of papers from the ACITY-2015, DPPR-2015, VLSI-2015, WiMNET-2015, AIAA-2015, CNDC-2015

We would like to thank the General and Program Chairs, organization staff, the members of the Technical Program Committees and external reviewers for their excellent and tireless work. We sincerely wish that all attendees benefited scientifically from the conference and wish them every success in their research. It is the humble wish of the conference organizers that the professional dialogue among the researchers, scientists, engineers, students and educators continues beyond the event and that the friendships and collaborations forged will linger and prosper for many years to come.

David C. Wyld
Natarajan Meghanathan

Organization

General Chair

Natarajan Meghanathan	Jackson State University, USA
Dhinaharan Nagamalai	Wireilla Net Solutions PTY LTD, Australia

Program Committee Members

Ahmed Hafaifa	University of Djelfa, Algeria
Abd El-Aziz Ahmed	Abd El-Aziz,Cairo University, Egypt
Abdelhafid Abdelmalek	Tlemcen University, Algeria
Abdelhamid Mansor	University of Khartoum, Sudan
Abdelouahab Moussaoui	Ferhat Abbas University, Algeria
Abdolreza Hatamlou	Islamic Azad University, Iran
Ahmad tayyar	ISRA University, Jordan
Ahmed Mohamed Khedr	University of Sharjah, UAE
Ajay Jaiswal	Rajiv Gandhi Technical University, India
Ali Chaabani	National School of Engineering, Tunisia
Anamika Ahirwar	Rajiv Gandhi Technical University, India
Andrews Samraj	Mahendra Engineering College, India
Arash Habibi lashkari	University of New Brunswick (UNB), Canada
Barbaros Preveze	Cankaya University Ankara, Turkey
Bela Genge	Petru Maior University of Tg. Mures, Romania
Chanabasayya Vastrad	Mangalore University, India
Chandramohan.D	Pondicherry University, India
Chun-Yi Tsai	National Taitung University,Taiwan
Dac-Nhuong Le	Haiphong University, Vietnam
Dean Wuzaa	Chang Jung Christian University, Taiwan
Emilio Jimenez	University of La Rioja, Spain
Faiyaz Ahmad	Integral University, India
Farhan	University Of Indonesia, Indonesia
Farshchi	Tehran University, Iran
Farzad Kiyani	Istanbul S.Zaim University, Turkey
Fatih Korkmaz	Cankiri Karatekin University, Turkiye
Foudil Cherif	Biskra University, Algeria
Girija Chetty	University of Canberra, Australia
Grienggrai Rajchakit	Maejo University, Thailand
Gullanar M Hadi	Salahaddin University, Hawler, Iraq
Hamdi M	National Engineering School of Tunis, Tunisia
Hossein Jadidoleslami	MUT University, Iran
Ian Tan	Multimedia University, Malaysia
Indumathi J	Anna University, India
Isa Maleki	Islamic Azad University, Iran
Israashaker Alani	Minustry If Science and Technolgy, Iraq
Jacques Epounde Ngalle	Robert Morris University, USA

Jan Zizka	Mendel University In Brno, Czech Republic
Jitendra Maan	Tata Consultancy Services, India
Jose Enrique Armendariz-Inigo	Public University of Navarre, Spain
Jyothi pillai	Bhilai Institute of Technology, India
Koushik Majumder	West Bengal University of Technology, India
Lakshmi Patibandla	Vignan's University, India
Mahesh K	Alagappa University, India
Majlinda Fetaji	South East European University, Macedonia
Maryam Kouzehgar	University of Tabriz, Iran
Meyyappan T	Alagappa University, India
Moez Hizem	Sup'Com, Tunisia
Mohamed Elboukhari	ESTO, Oujda, Morocco
Mohamed Hassan	American University of Sharjah, UAE
Mohamed Khamiss	Suez Canal University, Egypt
Mohammad Arif	Integral University, India
Mohammad Masdari	Islamic Azad University, Iran
Mucahit Altintas	Istanbul Technical University, Turkey
Muhammad Ali	University of Bradford, United Kingdom
Muhammad Sajjadur Rahim	University of Rajshahi, Bangladesh
Neela Narayanan V	VIT University, Chennai, India
Neetesh Saxena	State University of New York, South Korea
Nourddine Bouhmala	Buskerud and Vestfold University, Norway
Nourddine Bouhmala	Campus Vestfold, Norway
Peiman Mohammadi	Islamic Azad University, Iran
Polgar Zsolt Alfred	Technical University of Cluj Napoca, Romania
Prasad Halgaonkar	MIT College of Engineering, India
Raed I Hamed	University of Anbar Ramadi, Iraq
Rafah M. Almuttairi	University of Babylon, Iraq
Ram Kumar	SNS College of Engineering, India
Ramkumar Prabhu M	Dhaanish Ahmed College of Engineering, India
Reza Ebrahimi Atani	University of Guilan, Iran
Rim Haddad	Innov'com laboratory Sup'com, Tunisia
Saadat Pourmozafari	Tehran Poly Technique, Iran
Salem Nasri, Qassim University	Kingdom of Saudi Arabia
Seyyed AmirReza Abedini	Islamic Azad University, Iran
Shahid Siddiqui	Integral University, India
Smain Femmam	Uha University, France
Sokyna Qatawneh	Al-Zaytoonah University Of Jordan, Jordan
Soumen Kanrar	Vehere Interactive Calcutta, India
Suman Deb	NIT Agartala, India
Venkata Subramanian D	Hindustan University, India
Vitri Manar	Universitas Yarsi, Indonesia
William R Simpson	Institute for Defense Analyses, USA
Willie K Ofosu	Penn State Wilkes-Barre, USA
Zuhal Tanrikulu	Bogazici University, Turkey

Technically Sponsored by

Networks & Communications Community (NCC)

Computer Science & Information Technology Community (CSITC)

Digital Signal & Image Processing Community (DSIPC)

Organized By

Academy & Industry Research Collaboration Center (AIRCC)

TABLE OF CONTENTS

The Fifth International Conference on Advances in Computing and Information Technology (ACITY 2015)

The Fifth International Conference on Digital Image Processing and Pattern Recognition (DPPR 2015)

The Sixth International Conference on VLSI (VLSI 2015)

The Second International Conference on Wireless and Mobile Network (WiMNET 2015)

The Fifth International Conference on Artificial Intelligence, Soft Computing and Application (AIAA 2015)

The Second International Conference on Computer Networks & Data Communications (CNDC 2015)

DESIGNING A ROUTING PROTOCOL FOR UBIQUITOUS NETWORKS USING ECA SCHEME

Chandrashekhar Pomu Chavan and Pallapa Venkataram

Protocol Engineering and Technology Unit, Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore, India
{cpchavan, pallapa}@ece.iisc.ernet.in

ABSTRACT

We have designed a novel Event-Condition-Action (ECA) scheme based Ad hoc On-demand Distance Vector(ECA-AODV) routing protocol for a Ubiquitous Network (UbiNet). ECA-AODV is designed to make routing decision dynamically and quicker response to dynamic network conditions as and when event occur. ECA scheme essentially consists of three modules to make runtime routing decision quicker. First, event module receive event that occur in a UbiNet and split up event into event type and event attributes. Second, condition module obtain event details from event module split up each condition into condition attributes that matches event and fire the rule as soon as condition hold. Third, action module make runtime decisions based on event obtained and condition applied. We have simulated and tested the designed ECA scheme by considering ubiquitous museum environment as a case study with nodes range from 10 to 100. The simulation results show the time efficient with minimal operations.

KEYWORDS

Ubiquitous Network, Routing, Event, Subnet, Topology

1. INTRODUCTION

Ubiquitous Network (UbiNet) is a heterogeneous network [1] with various computing devices which are connected at anywhere, anytime and enable users to access and exchange information [2]. In UbiNet, the nodes may join/leave the network frequently and move freely hence, leads to frequent change in network topology. Since the nodes are mobile they can move arbitrarily in any direction leads to various failures like link failure, node failure and so on. Hence the dynamic network topology and frequent failures can be addressed using a routing protocol to manage the variety of failures and to choose optimal paths to transmit the data [3].

Routing in UbiNet is to find a best path from user to the service provider. Basically, routing algorithms are designed to determine the best paths in the network, whereas routing table entries store route information that the algorithm has already discovered a best path and routing protocols allow data packet to be collected and distributed across the network [4]. A UbiNet do

David C. Wyld et al. (Eds) : ACITY, DPPR, VLSI, WiMNET, AIAA, CNDC - 2015
pp. 01–18, 2015. © CS & IT-CSCP 2015 DOI : 10.5121/csit.2015.51301

not support any existing routing protocols since a client node does not have(or any node) or do not maintain the routing tables[5], [6].

We develop an ECA scheme based routing protocol in UbiNet i.e. ECA-AODV routing protocol for making dynamic ubiquitous routing decision quickly at runtime. In an ECA scheme, event [7] is defined as significant changes in state of a system. Event can occur at any time, event module collect the detail information about the event and forward to condition module for applying logical procedure in order to make dynamic routing decision quickly. When a specific event is occurred at a particular time, certain conditions are met then action module make dynamic decision [8],[9],[10].

1.1. Proposed Idea

We have proposed an ECA scheme for routing protocol in UbiNet. An ECA scheme is broadly divided into three modules namely Event module, Condition module and Action module respectively. Event module is a 2-tuples consists of event types and event attributes, the function of event module is to keep observe and notifies events. Condition module is 2-tuples consists of event details and condition attributes.

Condition module observes and receives an event, look for the rule that matches inputs and fire the rule as soon as condition hold. Based on event and condition, a decision is made by the action module which contains 1-tuple action attributes. ECA scheme is store using structure data structure and distributed across every node. Upon receiving event at a particular node, node intern broadcast occurred event to all its vicinity nodes.

1.2. Pattern of the Research Paper

Pattern of the research paper is as follows. Section 2 present most relevant works. Section 3 briefly explains AODV routing. Section 4 briefly describes routing in UbiNet. Section 5 explain proposed ECA scheme. Section 6 describes the ubiquitous museum environment case study. Section 7 gives the simulation environment. 8 shows the simulation result and finally, the paper draws some conclusions in Section 9.

2. SOME OF THE EXISTING WORKS

Bhandari S.R. and Bergmann N.W [11] describes program do not respond well for component or resource failure. An ECA based system is suitable for both describing desired system operations as well as linking an event [12] based system to communicate with resources.

Hannes Obweger et al.[13] propose an innovative framework for creating sense and response rules that can be useful for real-time. Sense and response has set of rules to detect event [14] from business stream and take appropriate decision based on event occurrence.

Kaan Bur and Cem Ersoy [15] presents a novel mesh based QoS multicast routing which keep track of the resources availability at every node and monitor the QoS status periodically. QoS multicast routing will be selected on the basis of available resources, protocol selects optimal route in heterogeneous network. Gateway node does protocol conversion at the boundaries of the subnet.

Jungil Heo and Wooshik Kim [16] analyze the information about user movement among heterogeneous subnet technologies in a ubiquitous system and proposes how to configure and manage the network in order to execute ubiquitous service in time. AODV protocol plays a vital role in health care environment to provide reliable and power efficiency data transmission.

Elizabeth M. Royer and Charles E. Perkins [17] consider fundamental functioning of AODV routing protocol in which, nodes store the route as and when needed. Each node uses destination sequence numbers to ensure freshness of the route at all times, AODV response quickly during link breakage in active routes.

3. AODV ROUTING

The AODV routing protocol is on-demand and reactive, in which path is discovered on-demand; established path is maintained as long as it is needed. AODV routing protocol [18] consists of four messages viz. i) Route request message, ii) Route reply message, iii) Route error message and iv) Route reply-acknowledgement message respectively. Essentially, node relies on intermediate nodes to find an optimal path from originating node to the ultimate target node in the network.

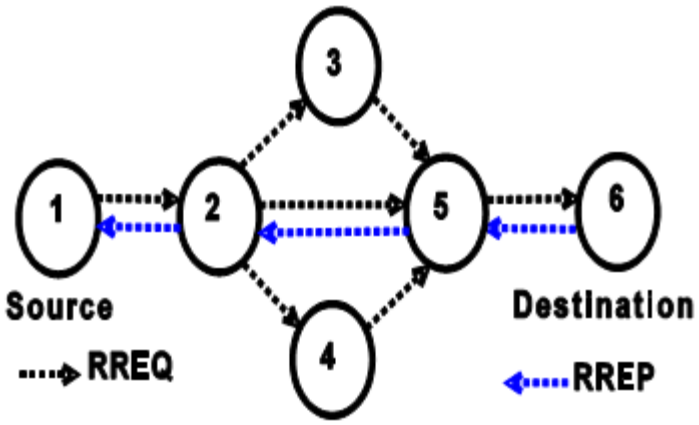


Figure 1. Route Discovery process in AODV Routing

In Figure 1, Node 1 has some data, it wishes to send to node 6 but node 1 does not have a valid path to communicate with node 6, in such case node 1 initiate path finding by sending route request message to its vicinity node 2, node 2 intern propagate the route request to its neighbour nodes such as 3, 4 and 5 respectively until route request reaches destination, if anyone of the intermediate node has the valid path toward the destination or node itself is destination may reply to the corresponding route request. Immediately upon receiving route request node verify that the sequence number [19] of the replying node is greater than that contained in route request message, then only node initiate to generate the route reply and unicast back to the route request originator. This is how a valid route is established in AODV routing.

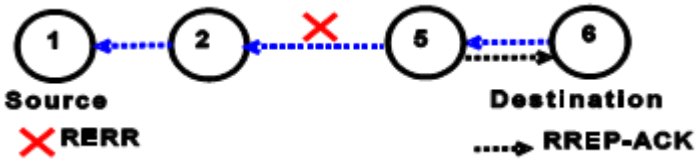


Figure 2. Route error and Route reply-ACK message

AODV uses route request and route reply for establishing a valid path, upon establishing valid path node store a routing table to communicate with rest of the nodes in the network, if existing path is not valid then node generate and forward the route error message to its predecessor node, when a node activate a bidirectional link then it send route reply acknowledgement message.

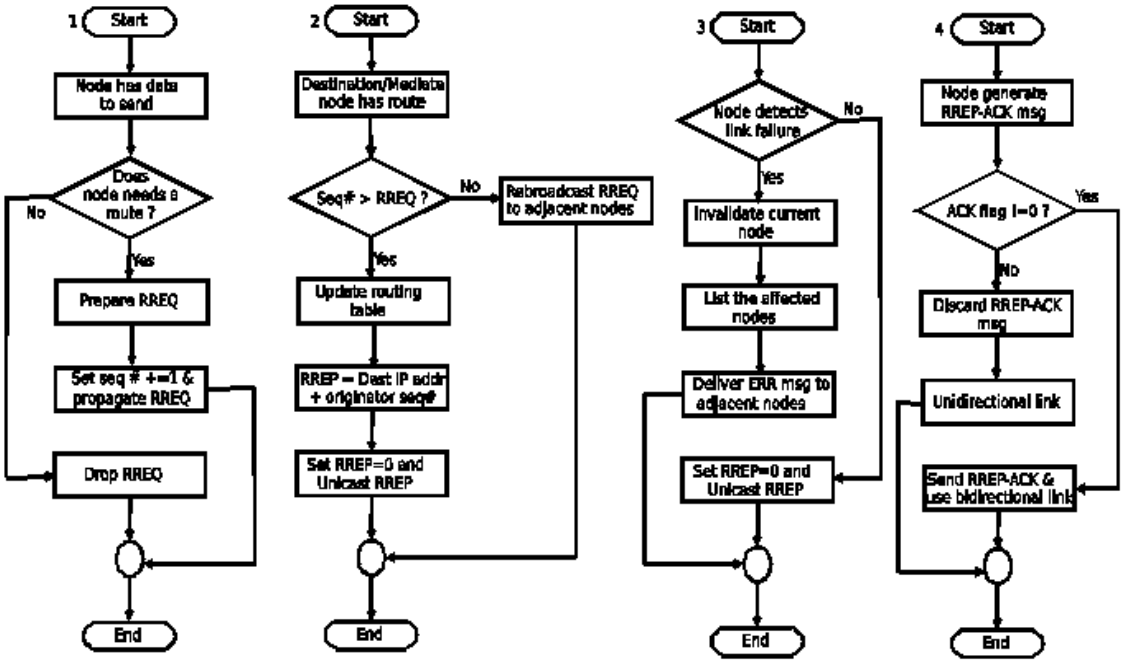


Figure 3. Flow chart of generating four AODV messages

Usually, in AODV originating node initiate a path discovery process as and when it would like to communicate with other node, node itself verify it has valid route or not if valid route does not exist then source node prepare RREQ packet and increment sequence number by one and propagate to its adjacent nodes.

4. ROUTING IN UBIQUITOUS NETWORKS

In ubiquitous environment, users demand constant availability of service at anytime from anywhere. Ubiquitous Server (UbiServ) receive informations about the user such as location, time, types of device, interest, preference, etc. from various embedded sensors and user profile, process the information and play a vital role in providing routing information to the user. UbiServ are connected and distributed across the internet, when a user enter into ubiquitous environment, an automatic path is establishes from user to the UbiServ without user's intervention [20].

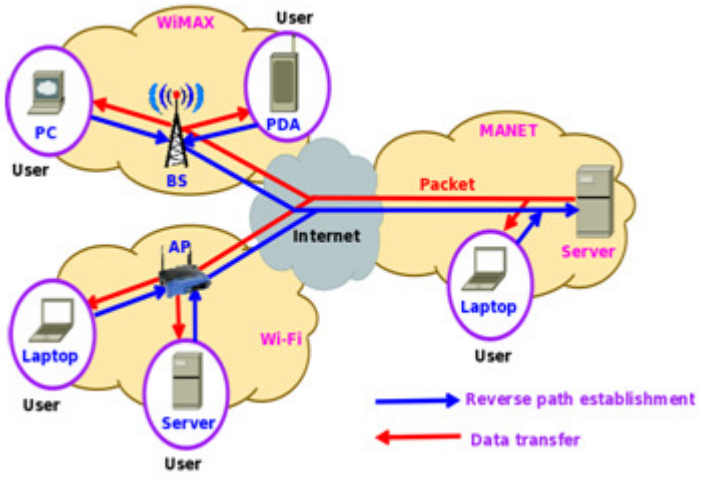


Figure 4. Example of routing in ubiquitous network

Routing in ubiquitous network has numerous differences as compared to normal routing in the following points

- Routing is adapted to heterogeneous network based on application requirements.
- UbiServ provide end-to-end flexible routing.
- Routing decision can be made dynamically based on current network status.
- Provide seamless service.
- Support QoS guaranteed service in high traffic.
- Lossless handover during switching from one network access technology to another network access technology.
- Provide best-effort traffic during congestion.

In ubiquitous network, a reverse path is established from user to the ubiquitous server, upon establishing path, UbiServ maintain an uninterrupted connectivity from user to the various subnet. UbiServ provide the routing information to the user located at heterogeneous subnet based on user's interest and preference.

5. PROPOSED ECA SCHEME IN UbiNet

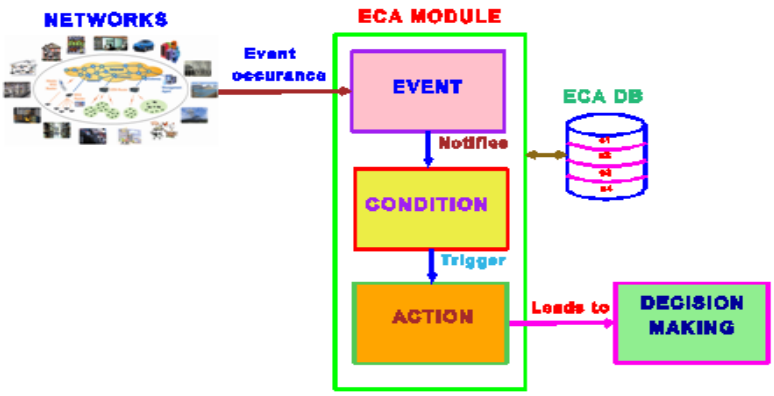


Figure 5. Function block diagram of an Event-Condition-Action Scheme

5.1 Function of Event Module

Event is defined as significant changes in state of a system. Event module keep observe events and notifies as and when events occurs. Event is a 2-tuple consists of event types and event attributes, event is denoted by E_i notation, event type may be time, spatial, composite, request, notification, internal, external, fault, service, etc.

$$E_i = (t_j, a_k) \quad (1)$$

Table 1. Event types

Event types (t_j)		
Type 1	Type 2 ...	Type m

Table 2. Event attribute

Event attributes (a_k)		
Data types	Parameters	Values

E_i is the i^{th} event, where $i \in \{1,2,3,...,n\}$, t_j is the j^{th} event type, a_k is the k^{th} event attribute, da_x is the data type, p_y is the parameter, v_z is attribute value

$$t_j \in \{t_1, t_2, t_3, ...,t_m\} \quad (2)$$

$$a_k=(da_x,p_y,v_z) \quad (3)$$

$$da_x \in \{da_1, da_2, da_3, ...,da_k\} \quad (4)$$

$$p_y \in \{p_1, p_2,p_3, ...,p_l\} \quad (5)$$

$$v_z \in \{v_1, v_2 ,v_3, ...,v_x\} \quad (6)$$

E_i occurs at a particular time is given by $E_i(t) = \begin{cases} 1; \text{Event has occured} \\ 0; \text{Otherwise} \end{cases}$

5.2 Function of Condition Module

Observe and receive an event, look for the rule that matches inputs and fire the rule as soon as condition hold. Condition module is 2-tuple consists of event details and condition attributes, $C_l=(d_x, c_m)$, Where C_l is the l^{th} condition, d_x is the x^{th} event details, c_m is the m^{th} condition attribute

$$d_x=(E_i,t_j,a_k) \quad \text{and } c_m=(c_p,a_q,o_r, r_t) \quad (7)$$

Table 3. Event details

Event details (d_x)
E_i, t_j, a_k

Table 4. Condition attributes

Condition attributes(c_m)			
Condition types	Arguments	Operators	Results

5.3. Function of Action Module

- Based on event and condition an operation to be carried out.
- Action is 1-tuple consists of action attributes.
- Action $A_n=(at_p)$, where A_n is the n^{th} action, at_p is the p^{th} action attribute.
- When $\{E_i\}$ occurs if $\{C_l$ true$\}$ then {Execute A_n }.
- ECA scheme is distributed across the network i.e $R:\{E_i,C_l\} \rightarrow A_n$.

Table 5. Action attributes

Action attributes (at_p)		
E_i	Condition result	Decision making

Algorithm1: Algorithm for dynamic routing decision using ECA scheme

1: Begin

2: **Input:** Set of events

3: **Output:** Dynamic routing decision

4: **if**(Event has occurred in Ubiquitous network) **then**

5: Split every event into event type and event attributes

6: Apply logical condition

7: Make runtime decision

8: **else**

9: Do not make dynamic routing decision

10: **end if**

11: End

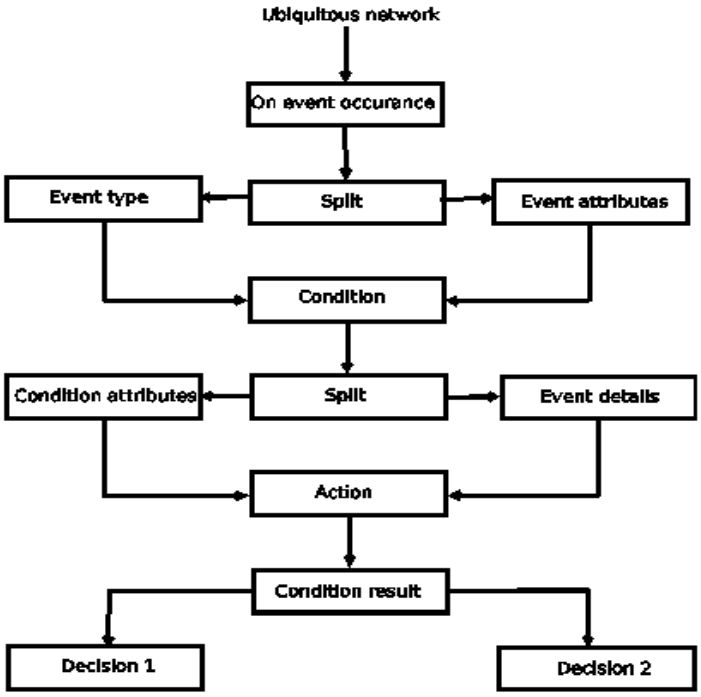


Figure 6. Event processing in ubiquitous network

5.4. ECA scheme based Routing in Ubiquitous Network

Table 6. Event(E1): Prepare route request and Event type(t1): Request

Event attributes(a_k)		
Data types	Parameters	Values
GUID	Event unique ID	9
datetime	Date&time of event occurred	10-03-2015 at 1:00pm
int	Packet type	RREQ=1
bool	Join flag	Set J=1 or 0
bool	Repair flag	Set R=1 or 0
bool	Gratuitous flag	Set G=1 or 0
bool	Destination only flag	Set D=1 or 0
int	Hop count	Initial value=0
int	Unique RREQ ID	4
string*ip_address	Destination IP address	10.32.21.1
int	Destination sequence #	13
string*ip_address	Source IP address	10.32.21.83
int	Source sequence number	5

Table 7. Condition(C1): Event attributes(d_x) are obtained from event module

Condition attributes(c_m)			
Condition types	Arguments	Operators	Results
Condition1: Checking valid path	Status	?:	No valid path exist toward destination node
Condition2: Prepare RREQ message and setting sequence#	Set	+=	Prepare RREQ message & set sequence number +=1

Table 8. Action(A1): Event and condition details are obtained

Action attributes(at_p)		
Event	Condition result	Decision making
Prepare route request	Node does not have route to destination &&Prepare RREQ and set seq # += 1	RREQ message is prepared and ready to broadcast

ECA Rule for Event(E1) : When{Prepare route request event occurs} **If**{(Node does not have valid route) && (Sequence number+= 1)} **Then {**RREQ message is prepared and ready to broadcast}

Table 9. Event(E2):Generate route reply and Event type(t2): Request

Event attributes(a_k)		
Data types	Parameters	Values
GUID	Event unique ID	23
datetime	Date&time of event occurred	11-04-2015 at 2:00pm
int	Packet type	RREP=2
bool	Repair flag	Set R=1 or 0
bool	Acknowledgement flag	Set A=1 or 0
bool	Prefix size	PS=00000
int	Hop count	Initial value=0
string*ip_address	Destination IP address	10.32.21.83
int	Destination sequence #	2
string*ip_address	Source IP address	10.32.21.1
time_t	Lifetime	25msec

Table 10. Condition(C2): Event attributes(d_x) are obtained from event module

Condition attributes(c_m)			
Condition types	Arguments	Operators	Results
Condition1:Destination node	Status	!=	Destination node that has active route
Condition2:Sequence number	Verify	>	Sequence number must be greater than that contained in RREQ message

Table 11. Action(A2): Event and condition details are obtained

Action attributes(at_p)		
Event	Condition result	Decision making
Generate route reply	Destination node that has active route && sequence # is greater than RREQ	Reverse route reply to source node

ECA Rule for Event(E2): When{Generate route reply event occurs} **If**{(Destination node that has active route) && (Sequence# > RREQ message)} **Then** {Reverse route reply to source node}

Table 12. Event(E3):Route link has been broken and Event type(t3): Notification

Event attributes(a_k)		
Data types	Parameters	Values
GUID	Event unique ID	25
datetime	Date&time of event occurred	14-04-2015 at 5:00pm
int	Packet type	RERR=3
bool	No delete flag	Set N=1 or 0
int	Destination count	Initial value=0
string*ip_address	Destination IP address	10.32.21.51
int	Destination sequence #	6

Table 13. Condition(C3): Event attributes(d_x) are obtained from event module

Condition attributes(c_m)			
Condition type	Argument	Operator	Result
Condition1:Link fail	Check	?:	Invalidate the route

Table 14. Action(A3): Event and condition details are obtained

Action attributes(at_p)		
Event	Condition result	Decision making
Route link has been broken	Invalidate the route	List the affected nodes

ECA Rule for Event(E3): When{Route link has been broken event occurs} **If**{Invalidate the route} **Then {** List the affected nodes}

Table 15. Event(E4):Generating route reply-ack and Event type(t4): Request

Event attributes(a_k)		
Data types	Parameters	Values
GUID	Event unique ID	14
datetime	Date&time of event occurred	21-04-2015 at 2:00pm
int	Packet type	RREP-ACK=4
bool	Repair flag	Set R=1 or 0
bool	Acknowledgement flag	Set A=1 or 0
bool	Prefix size	PS=00000
int	Hop count	Initial value=0
string*ip_address	Destination IP address	10.32.21.88
int	Destination sequence #	7
string*ip_address	Source IP address	10.32.21.13
time_t	Lifetime	25msec

Table 16. Condition(C4): Event attributes(d_x) are obtained from event module

Condition attributes(c_m)			
Condition type	Argument	Operator	Result
Condition1:Acknowledgement flag	Set	=	Use bidirectional link if A=1,

Table 17. Action(A4): Event and condition details are obtained

Action attributes(at_p)		
Event	Condition result	Decision making
Generating route reply-ack	Use bidirectional link if A=1,	Send route reply-ack

ECA Rule for Event(E4): When{Generating route reply-ack} **If**{Bidirectional link flag A=1} **Then {** List the affected nodes}

5.5. ECA state transition diagram

- Event transition changes from one state to another state when initiated by a triggering event or condition.
- State machine is given by pentuple i.e. SM =($\sum$, S, s_0, δ, F).
- $\sum$ is the input which is considered as set of events $\sum$= {E_i}.
- S is the #of states i.e. state of a node S={1,2,...,n}.
- s_0 is the initial state, where an event has occurred.
- δ be the transition which is considered as condition.
- δ = {$State_{(old)} \rightarrow State_{(new)}$,$Input_{(condition)} \rightarrow Output_{(condition)}$ }.
- F is the final state, where an action will be taken.

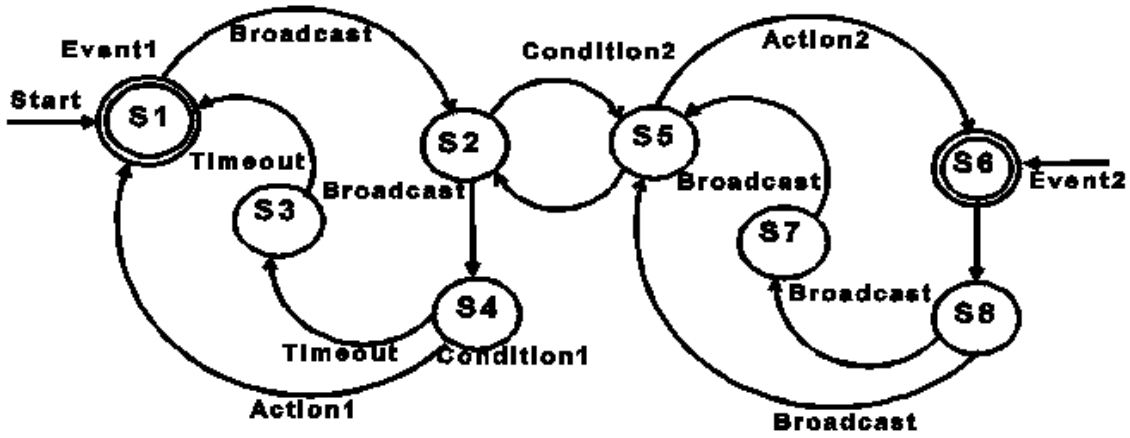


Figure 7. State transition diagram of an ECA Scheme

6. CASE STUDY : ECA SCHEME IN UBIQUITOUS MUSEUM ENVIRONMENT

Table 18. ECA scheme in ubiquitous museum environment

Event-Condition-Action		
Event	Condition	Action
User is looking for a route to visit exhibit	(Interest=Science) &&(Preference = Biology)	Provide route information about biology exhibit
Lunch time	(Preference=North-Indian) && (Time > 1 PM)	Route to restaurant
High temperature in museum	Temperature $\geq 30^0$ C	Switch on AC
User blood pressure is low	BP < 90/60	Provide shortest route to hospital
User is spending more time in front of exhibit	User's history says user is new to museum	Provide details information about exhibit

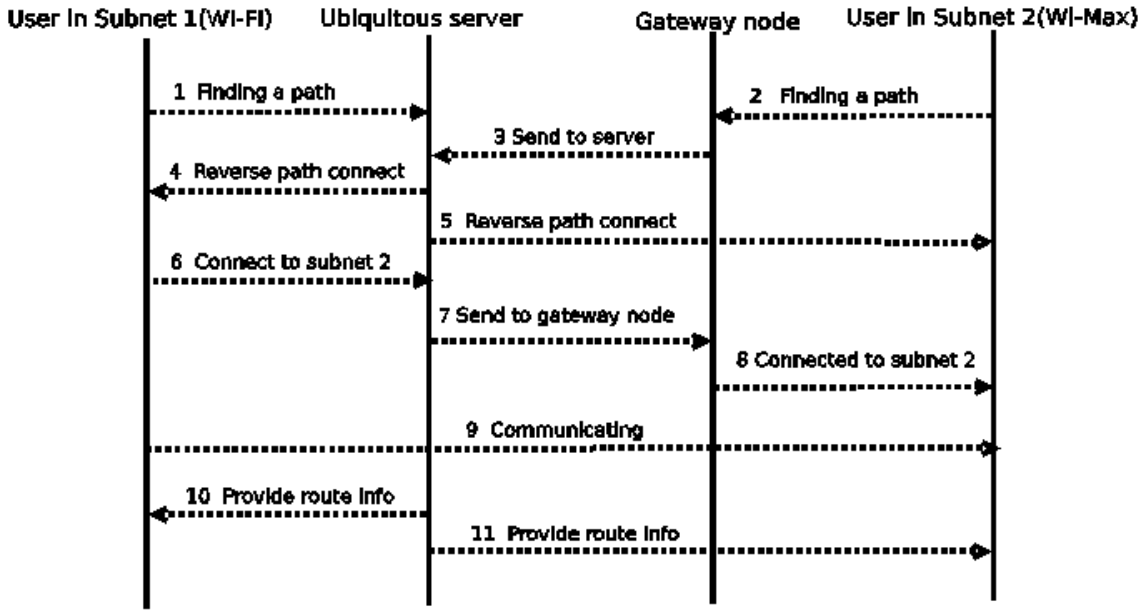


Figure 8. Sequence diagram to establish path between user and ubiquitous server

7. SIMULATION ENVIRONMENT

Proposed ECA scheme is implemented using C programming language and simulated in NS 2.34 simulator, by considering simulation parameters as depicted in table below. We have consider two different subnets such as Wi-Fi, MANETs in which random nodes are created and reverse path is established from ubiquitous server to the ubiquitous user, routing information is provided to the individual user based on interest, preferences.

Table 19. Simulation parameters

Simulation parameters	
Parameters	Values
Nodes	100
Routing protocol	ECA-AODV
Transmission range	30mts
Simulation time	500s
Topology size	25mX25m
Packet size	512 Bytes
Mobility	Random

8. SIMULATION RESULTS

ECA scheme is simulated using NS 2.34 simulator and set of results are obtained as shown in the following figures. In fig #9, we have created 50 random nodes and simulated for 500ms for AODV and ECA-AODV protocol, ECA-AODV protocol achieve lower packet delivery latency than normal AODV protocol. Fig #10 in which 100 random nodes are created, different events are consider as an input for simulation by using 2 different routing protocols such as AODV and ECA-AODV respectively, after result execution, we conclude that data packet sending ratio of ECA-AODV is good in comparison with conventional AODV as and when event occur in the UbiNet.

In fig #11, we have shown mobility speed versus control byte transferred over data byte delivered with normal routing and ECA based routing. Fig #12 describes ECA scheme based RREQ and RREP message speed is high i.e. quicker response as and when event occur than the normal route request, route reply message and finally, fig #13 explain number of event processed per second by the node in ECA-AODV is better as and when number of node increases compared to the conventional AODV routing protocol.

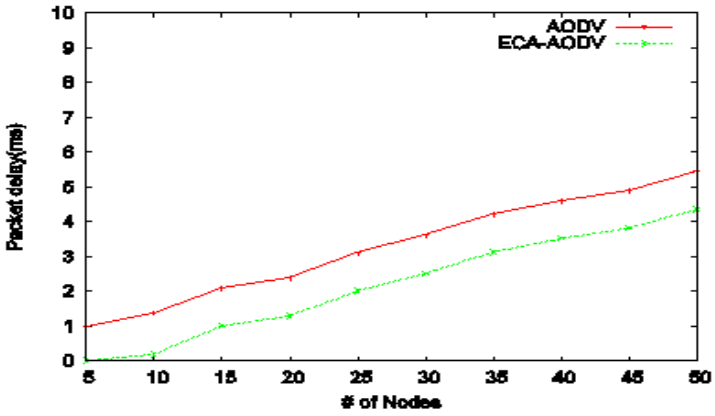


Figure 9. #of nodes in network Vs packet delivery

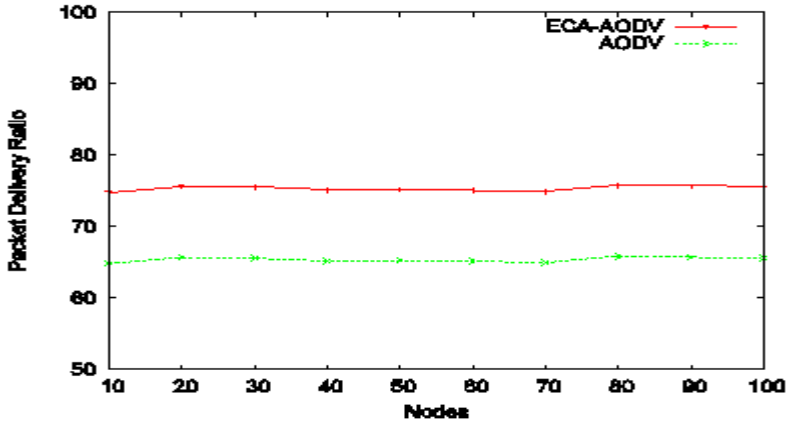


Figure 10. #of nodes Vs data packet delivery ratio

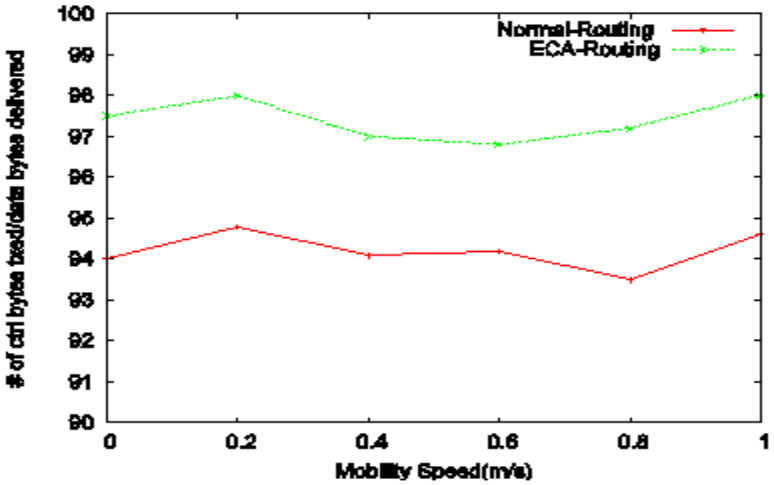


Figure 11. Mobility speed Vs #of ctrl bytes transferred/data bytes delivered

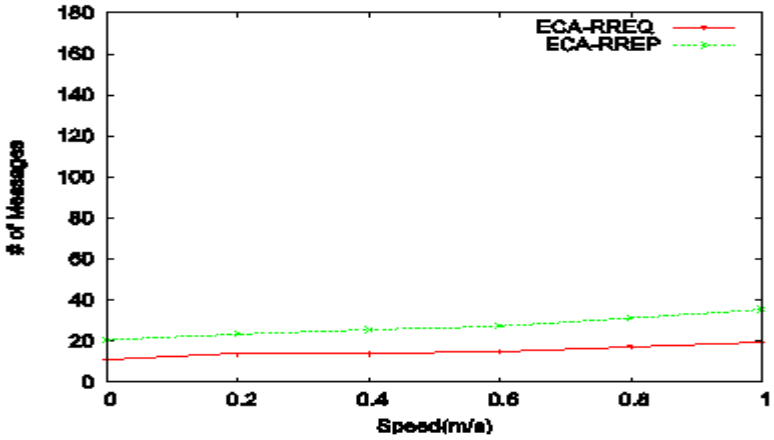


Figure 12. Speed Vs #of messages

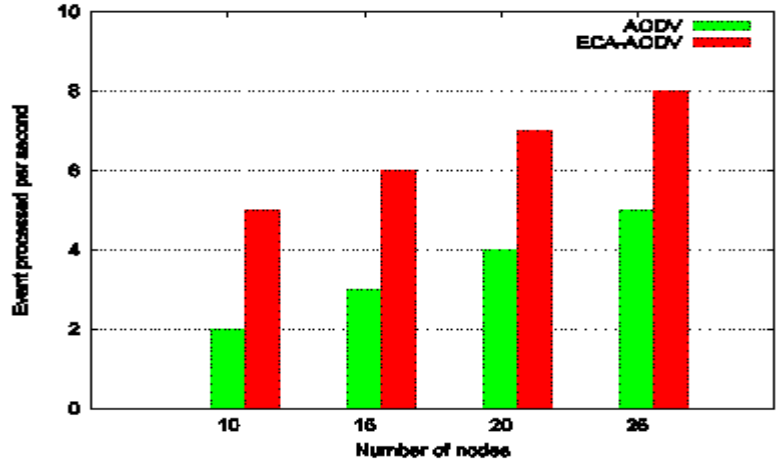


Figure 13. #of nodes Vs event processed per second

9. CONCLUSIONS

We have designed and simulated routing protocol using a novel ECA scheme in a UbiNet. Event module is designed using 2-tuples consists of event types and event attributes, condition module is designed using 2-tuples consists of event details and condition attributes and finally, an action module is designed using 1-tuple. We have considered ubiquitous museum environment as a case study to show the simulation of our proposed scheme, in which ubiquitous server provide uninterrupted connectivity and routing related information to the user as and when user move from one subnet to another subnet. However the proposed ECA scheme is flexible to supports dynamic network conditions in heterogeneous subnet and hence scheme is more effective as well as efficient as compared to non-ECA scheme in terms of parameters such as flexibility during runtime, easily adapt to the network dynamicity, quicker response as soon as event occur and easily adaptable to the types of network access technology used.

ACKNOWLEDGMENTS

We would like to thank our colleagues at Protocol Engineering and Technology unit, Department of ECE, Indian Institute of Science, Bangalore, India for their help and anonymous reviewers for their constructive and most valuable suggestions on improving the quality of the paper.

REFERENCES

[1] Chawathe, Y.; McCanne, S.; Brewer, E.A., "RMX: reliable multicast for heterogeneous networks," INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE, vol.2, no., pp.795, 804 vol.2, 2000.

[2] T. Murakami and A. Fujinuma. Ubiquitous networking towards a new paradigm, Nomura research institute. April 2000.

3] Wang, H., Zhang, Y., & Cao, J. (2006, May). Ubiquitous computing environments and its usage access control. In Proceedings of the 1st international conference on Scalable information systems (p. 6). ACM.

[4] Ge, M., Krishnamurthy, S. V., and Faloutsos, M. (2006). Application versus network layer multicasting in ad hoc networks: the ALMA routing protocol. Ad Hoc Networks, 4(2), 283-300.

[5] Young - Guk Ha; Joo - Chan Sohn; Young - Jo Cho, "ubiHome: An Infrastructure for Ubiquitous Home Network Services," Consumer Electronics, 2007. ISCE 2007. IEEE International Symposium on, vol., no., pp.1,6, 20-23 June 2007.

[6] Lee, Sung-Ju, William Su, and Mario Gerla. "On-demand multicast routing protocol in multi-hop wireless mobile networks." Mobile Networks and Applications 7.6 (2002): 441- 453.

[7] Orriëns, Bart, Jian Yang, and Mike P. Papazoglou. "Model driven service composition." Service-Oriented Computing -ICSOC 2003. Springer, Berlin Heidelberg, 2003. 75- 90.

[8] Yin, Liuguo, Changmian Wang, and Geir E. Øien. “An energy - efficient routing protocol for event driven dense wireless sensor networks." International Journal of Wireless Information Networks 16.3 (2009): 154-164

[9] M Weiser. The computer of the 21st century. In ACM SIGMOBILE. Mobile computing and Communications. Pages 3-11, July 1999

[10] Levina, Olga, and Vladimir Stantchev. "A Model and an Implementation Approach for Event-Driven Service Orientation." International Journal on Advances in Software 2.2 and 3 (2009): 288-299

[11] Bhandari, Shiddartha Raj, and Neil W. Bergmann. "An internet-of-things system architecture based on services and events." Proceedings of the 2013 IEEE 8th International Conference on Intelligent Sensors, Sensor Networks and Information Processing: Sensing the Future, ISSNIP 2013. Vol. 1. IEEE, 2013.

[12] Ye, Yan, Zhibin Jiang, Xiaodi Diao, and Gang Du. "Extended event-condition -action rules and fuzzy Petri nets based exception handling for workflow management." Expert Systems with Applications 38, no. 9 (2011): 10847-10861.

[13] Obweger, Hannes, Josef Schiefer, Martin Suntinger, and Peter Kepplinger. "Model-driven rule composition for event-based systems." International Journal of Business Process Integration and Management 5, no. 4 (2011): 344-357.

[14] Alferes, José Júlio, Federico Banti, and Antonio Brogi. "An event-condition-action logic programming language." Logics in Artificial Intelligence. Springer Berlin, Heidelberg, 2006. 29- 42.

[15] Bur, K.; Ersoy, C., "Quality-of-service-aware multicast routing in heterogeneous networks with Ad hoc extensions," Computer and Information Sciences, 2008. ISCIS '08. 23rd International Symposium on , vol., no., pp.1,6, 27-29 Oct. 2008

[16] Heo, Jungil, and Wooshik Kim. "Ad-hoc routing Protocol-based Ubiquitous Network System In the Hospital Environment." In Antennas, Propagation & EM Theory, 2006. ISAPE'06. 7th International Symposium on, pp. 1-4. IEEE, 2006.

[17] C. E.Perkins and E. M.Royer. The ad hoc on-demand distance vector routing protocol. In Charles E.Perkins editor, Ad hoc networking Addison-Wesley, 2000

[18] C. Perkins, E. Belding-Royer, and S. Das. Ad hoc on-demand distance vector routing protocol in www.ietf.org/rfc/rfc3561, July 2003

[19] Zarei, M., Faez, K., & Nya, J. M. (2008, December). Modified Reverse AODV routing algorithm using route stability in mobile ad hoc networks. In Multitopic Conference,2008. INMIC 2008. IEEE International (pp. 255-259).

[20] P. Venkataram and M. Bharath. Context based service discovery for ubiquitous applications. Information Networking(ICOIN), 2011 International conference on , pages 311-316, Jan 2011

AUTHORS

Chandrashekhar Pomu Chavan received his BE in Computer Science and Engineering from Guru Nanak Dev Engineering College, Bidar, Karnataka, India, and M.Tech Degree in Network and Internet Engineering from Sri Jayachamarajendra College of Engineering, Mysore, Karnataka, India, in 2005 and 2008 respectively. Currently he is pursuing his Ph.D degree on Ubiquitous Network under the guidance of Prof. Pallapa Venkataram in the Department of Electrical Communication Engineering at Indian Institute of Science, Bangalore, India. His research interests are in the areas of Ubiquitous Computing, Pervasive Computing, Mobile Ad hoc Network, Context Aware System, and Routing Protocols.

Prof. PallapaVenkataram received his Ph.D. Degree in Information Sciences from the University of Sheffield, England, in 1986. He is currently the chairman for center for continuing education, and also a Professor in the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore, India. Dr. Pallapa's research interests are in the areas of Wireless Ubiquitous Networks, Social Networks, Communication Protocols, Computation Intelligence applications in Communication Networks and Multimedia Systems. He is the holder of a Distinguished Visitor Diploma from the Orrego University, Trujillo, PERU. He has published over 200 papers in International/national Journals/conferences. Written three books: Mobile and wireless application security, Tata McGraw-Hill, Communication Protocol Engineering, Prentice-Hall of India (PHI), New Delhi, 2004 (Co-author: Sunil Manvi), and Multimedia: Concepts & Communication, Darling Kinderley(India) Pvt. Ltd., licensees of Pearson Education in South Asia, 2006. Edited two books: Wireless Communications for Next Millennium, McGraw-Hill, 1998, and Mobile Wireless Networks & Integrated Services, John Wiley & Sons(Asia) Pvt. Ltd., 2006(Co-editors: L.M.Patnaik & Sajal K. Das). Written chapters for two different books, and a guest editor to the IISc Journal for a special issue on Multimedia Wireless Networks. He has received best paper awards at GLOBECOM'93 and INM'95 and also CDIL (Communication Devices India Ltd) for a paper published in IETE Journal. He is a Fellow of IEE (England), Fellow of IETE (India) and a senior member of IEEE Computer Society.

TOPIC MODELING: CLUSTERING OF DEEP WEBPAGES

Muhunthaadithya C[1], Rohit J.V[2], Sadhana Kesavan[3] and
Dr. E. Sivasankar[4]

[1, 2, 3]Department of CSE, NIT Trichy
4Assistant Professor, Department of CSE, NIT Trichy-620015,
Tamilnadu, India
muhunth93@gmail.com
jvrohit1201@gmail.com
sadhanakesavan@gmail.com

ABSTRACT

The internet is comprised of massive amount of information in the form of zillions of web pages.This information can be categorized into the surface web and the deep web. The existing search engines can effectively make use of surface web information.But the deep web remains unexploited yet. Machine learning techniques have been commonly employed to access deep web content.

Under Machine Learning, topic models provide a simple way to analyze large volumes of unlabeled text. A "topic" consists of a cluster of words that frequently occur together. Using contextual clues, topic models can connect words with similar meanings and distinguish between words with multiple meanings. Clustering is one of the key solutions to organize the deep web databases.In this paper, we cluster deep web databases based on the relevance found among deep web forms by employing a generative probabilistic model called Latent Dirichlet Allocation(LDA) for modeling content representative of deep web databases. This is implemented after preprocessing the set of web pages to extract page contents and form contents.Further, we contrive the distribution of "topics per document" and "words per topic" using the technique of Gibbs sampling. Experimental results show that the proposed method clearly outperforms the existing clustering methods.

KEYWORDS

Latent Dirichlet Allocation, Latent Semantic Analysis, Deep Web, Cosine Similarity, Form Content and Page Content.

1. INTRODUCTION

1.1 Deep Web

The internet is a huge repository for an extremely wide range of information. Most of us access this information by querying through standard search engines. However, a portion of this World

David C. Wyld et al. (Eds) : ACITY, DPPR, VLSI, WiMNET, AIAA, CNDC - 2015
pp. 19–27, 2015. © CS & IT-CSCP 2015 DOI : 10.5121/csit.2015.51302

Wide Web content is not explored by any standard search engines[1]. This content is "masked" and is referred to as the deep web[2].

The deep web is significantly gigantic. According to a July 2000 white paper [3], the deep web is 500 times larger than the surface web and indeed, continues to proliferate this magnitude[3].From the surveys presented in [4],[5] we can get a lucid understanding of the surface web. There are links buried far down on sites thriving on the surface web, which direct us to the news and history of the Deep web. It is quite common to come across deep web pages that have links terminating with the extension '.onion'. Such web pages require us to access them through browsers named 'Tor', which connect to the servers containing the repository and get access consent[6].

Today, the 60 largest Deep Web sites contain around 750 terabytes of data, surpassing the size of the entire Surface Web 40 times. 95% of the Deep Web is publically accessible, which is free of cost. This is why search engines, such as Google, index well over a trillion pages on the World Wide Web, but there is information on the web that common search engines don't explore. Most of this constitutes databases of information that need to be searched directly from the specific website. A small pocket of the deep web is filled with hyper-secret communities who flock there to escape identification from authorities [7].

1.2 Topic Modeling

Topic models provide a simple way to analyze large volumes of unlabeled text. A topic consists of a cluster of words that frequently occur together. Using contextual clues, topic models can connect words with similar meanings and distinguish between uses of words with multiple meanings. Topic models express the semantic information of words and documents by 'topics'[8].

1.1 Clustering

Clustering is a division of data into groups of similar objects where each group consists of objects that are similar between themselves and dissimilar to objects of other groups.

From the machine learning perspective, clustering can be viewed as unsupervised learning of concepts. We can cluster images, patterns, shopping items, feet, words, documents and many more fields. Clustering has wide scope of application in data mining, text mining, information retrieval, statistical computational linguistics, and corpus based computational lexicography.

Clustering has the following advantages**:**

1. Clustering improves precision and recall in information retrieval.
2. It organizes the results provided by search engines.
3. It generates document taxonomies.
4. It also generates ontologies and helps in classifying collocations.

A good clustering algorithm needs to have the characteristics, mentioned below:

1. Scalability
2. High dimensionality

3. Ability to discover clusters of arbitrary shape

Considering the above factors and enormous size of documents, we have analyzed that Latent Dirichlet allocation outperforms the efficiency of other existing clustering algorithms [9].

2. RELATED WORK

Several deep web clustering approaches exist. Here, we briefly discuss the motivation for our work. The content in [11] proposes to cluster deep web pages by extracting the page content and the form content of a deep web page. Considering the extracted parts as words in a document, clustering is done. The results prove to be better than those of several previous models. However, for some datasets, words were clustered under irrelevant topics.

3. PROPOSED METHOD

A topic consists of a cluster of words that frequently occur together. Topic models can connect words with similar meanings and distinguish between uses of words with multiple meanings.A variety of topic models have been used to analyze the content of documents and classify the documents. For example, Latent Semantic Analysis compares documents by representing them as vectors, and computing their dot product. In our paper, we have used another such algorithm: Latent Dirichlet Allocation [10].

3.1 Latent Dirichlet Allocation

Latent Dirichlet Allocation (LDA) is a generative probabilistic topic model for collections of discrete data. The generative model describes that the documents are generated using the following algorithm. For each document:

1. Firstly, a distribution over topics is chosen randomly.
2. Now, for each word to be generated in the document,
   a. A topic is chosen from the distribution created in Step 1.
   b. Next, a word is chosen randomly from the corresponding distribution over vocabulary for the topic.

The Dirichlet prior on per document topic distribution is given by

$$p(\theta|\alpha) = \frac{\Gamma(\sum_{i=1}^{k} \alpha_i)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \theta_1^{\alpha_i - 1} \ldots . \theta_k^{\alpha_k - 1} \quad (1)$$

Joint distribution of topic mixture θ, a set of N topic z, a set of N words w

$$p(\theta, z, w|\alpha, \beta) = p(\theta|\alpha) \prod_{n=1}^{N} p(z_n|\theta) p(w_n|z_n, \beta) \quad (2)$$

α -hyper parameter on the mixing proportions (K-vector or scalar if symmetric).

β -hyper parameter on the mixture components (V-vector or scalar if symmetric).

Θ_m-parameter notation for p(z|d=m), the topic mixture proportion for document m. One proportion for each document, $\Theta = \{\vartheta\sim m\}$ M m=1 (M × K matrix).

Φ_k-parameter notation for p(t|z=k), the mixture component of topic k. One component for each topic, $\Phi = \{\phi\sim k\}$ K k=1 (K × V matrix).

N_m- Document length (document-specific), here modeled with a Poisson distribution with constant parameter ξ.

$Z_{m,n}$- Mixture indicator that chooses the topic for the nth word in document m.

$W_{m,n}$ - Term indicator for the nth word in document m.

In order to retrieve topics in a corpus, this generative process is reversed. The two distributions: distribution over topics and distribution over vocabulary for the topics are discovered by the reverse process. This process uses Gibbs Sampling, which is commonly used as a means of statistical inference. It is a randomized algorithm (i.e. an algorithm that makes use of random numbers, and hence may produce different results each time it is run). Gibbs sampling generates a Markov chain of samples, each of which is correlated with nearby samples.

3.2 Parsing

LDA is a document clustering algorithm. To give input to LDA as documents, we need to parse the input XML/HTML files. As mentioned earlier, the Page Contents and the Form Contents are extracted from the web pages.

1. Page Content: Values and textual content visible on the web page.
2. Form Content: The value of the attributes in the form.

Most of the deep web pages are blocked by forms (e.g. login forms). To take advantage of this, we take into consideration the form attribute values, which can reveal important information about the type of deep web pages. Thus, this approach covers a large portion of deep web database for clustering.

3.3 Visiting Inner Links

For some web pages, it is possible that there is very little information related to the topic it comes under. However, the webpage may contain links to other related sites, or even its home page. If we were able to visit the inner HTML links in the webpage, we could gather further information about the topic the webpage talks about. For this purpose, we perform parsing individually on all the links inside the webpage. This way, we also gather more relevant test data than what is already available.

However, is also possible to have irrelevant links inside the webpage. For example, there might be a quick link to a search engine, or advertisements. Since the web pages discovered by visiting the inner links are not so reliable, we must give them lesser weightage than the original data. We achieve this by increasing the frequency of the words in the original webpage, i.e., we repeat the words in the original webpage for a specific number of times.

3.4 Gibbs sampling

Gibbs sampling is one of the class of sampling methods known as Markov Chain Monte Carlo. We use it to sample from the posterior distribution, p(Z|W), given the training data W represented in the form of

$$W = \begin{bmatrix} \{w_1, \ldots, w_{N_1}\}, & \text{words in the 1st document} \\ \{w_{N_1+1}, \ldots, w_{N_1+N_2}\}, & \text{words in the 2nd document} \\ \ldots & \ldots \\ \{w_{1+\sum_{j=1}^{D-1} N_j}, \ldots, w_{\sum_{j=1}^{D} N_j}\} & \text{words in the } D\text{-th document} \end{bmatrix},$$

Gibbs sampling is commonly used as a means of statistical inference, especially Bayesian inference. It is a randomized algorithm (i.e. an algorithm that makes use of random numbers, and hence may produce different results each time it is run), and is an alternative to deterministic algorithms for statistical inference such as variational Bayes or the expectation-maximization algorithm (EM). Gibbs sampling generates a Markov chain of samples, each of which is correlated with nearby samples.

3. IMPLEMENTATION

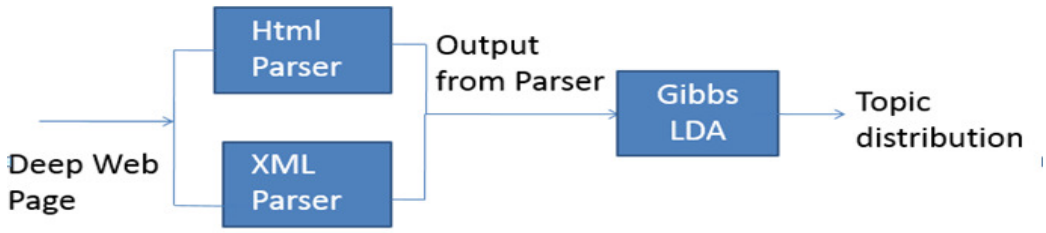


Fig 2: Module Diagram

We split the implementation into two parts:

1. Parsing
2. LDA

4.1 Parsing

Input: Deep web interface webpage (HTML/XML)

Output: Document with Form Contents and Page Contents as words.

4.1.1 Algorithm

1. From the input HTML/XML, gather all the textual content. Textual content refers to all the *text* that is visible on the webpage.
2. For each <form> element, gather all the values of the attributes for each of the child tags. For example, in the following we wish to extract the *"Search by Departure City"* and *"departure city"*.

<form>
<attrgroup>
<attr name="Search by Departure City" ename="departure city">
…
…
</attrgroup>
</form>

3. The words extracted from *Step 1* and *Step 2* constitute the *extracted words.* Write the *extracted words* into the output document.
    a. If the webpage is the original webpage interface, repeat the extracted word for a fixed number of times (say N), and write to the output document.
    b. If the webpage is among the ones discovered by visiting the inner links, write them to the document as it is.
4. Remove all the stop-words. Stop-words constitute certain regularly used words which do not convey any meaning, like *as, the, of, a, on,*etc.

4.1.2 Output Format

Google App Engine has been used for creating the web-tool for this project. The parser code written in Python takes in multiple HTML/XML files for input and gives the output after extracting the required words. Further, in order to remove grammatical constructs like ‘a’, ‘the’, ‘on’, we include a file of *stop-words*. All the words in this file are checked against the words obtained from the App Engine (Python code), and removed.

The input format for LDA is as follows:
<Number of documents: N>
<Document 1 - space separated words>
<Document 2 - space separated words>
…
…
<Document N - space separated words>
This output file from the python parser code produces the output in the above mentioned format. This output file is then given as input to the LDA code.

4.2 Gibbs LDA

Input:

1. A single file constituting the set of documents received after the Parsing stage. Multiple web pages parsed and given as input to LDA.

2. Number of Topics to be discovered.

Output: Clusters - A set of topics with most occurring words in each topic.

4.2.1Algorithm

As described earlier, we feed the Gibbs sampled data to our LDA model to extract topics from the given corpus.

5. EXPERIMENTAL RESULTS

In order to evaluate the performance of our approach we tested it over dataset of TEL-8 as mentioned in [8], UIUC Web integration repository [12]. The repository contains web search forms of 447 sources originally classified into 8 domains. The term TEL-8 refers to eight different web source domains belonging to three major categories (Travel, Entertainment and Living). Travel group is related to car rentals, hotels and airfares. Entertainment group has books, movies and music records interfaces. Finally the living group contains jobs and automobiles related query interfaces.

5.1 Performance measure

To evaluate the performance of our method, we calculated Precision of the outputs generated.In a classification task, the precision for a class is the number of true positives divided by the total number of elements labeled as belonging to the positive.

The definition of related terms is shown below:

1. TN / True Negative: case was negative and predicted negative
2. FP / False Positive: case was negative but predicted positive

[13] Precision = True Positives / (True positives + False Positives)

We tried various combinations of these 8 domains, and created 4 small datasets and 2 large datasets. We compared the approach used in [8] and our approach over the same dataset, and found that our method performs well in most cases.

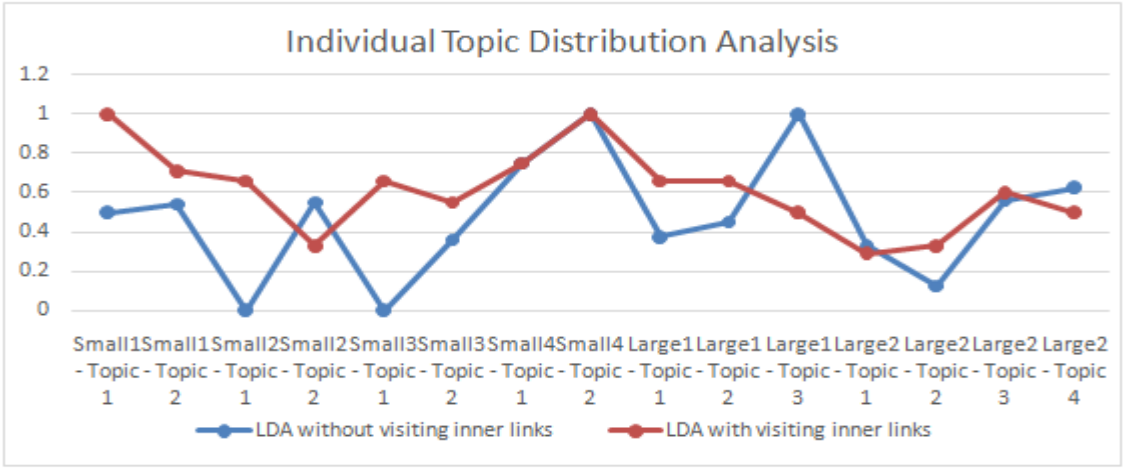


Fig 3: Individual Topic Distribution Analysis. Small refers to small dataset, and Large refers to large dataset. Several topics are discovered within for each dataset.

Though, this analysis clearly shows that our method performs better, this does not give the complete picture. For a better overall picture of the total topic distribution in each of the datasets, we use overall precision for each dataset.

We use the same formula for precision used above, and average it for the over the number of topics. Therefore, we get an averaged precision value for each dataset.

Averaged Precision = Sum of all precision values / Number of topics discovered

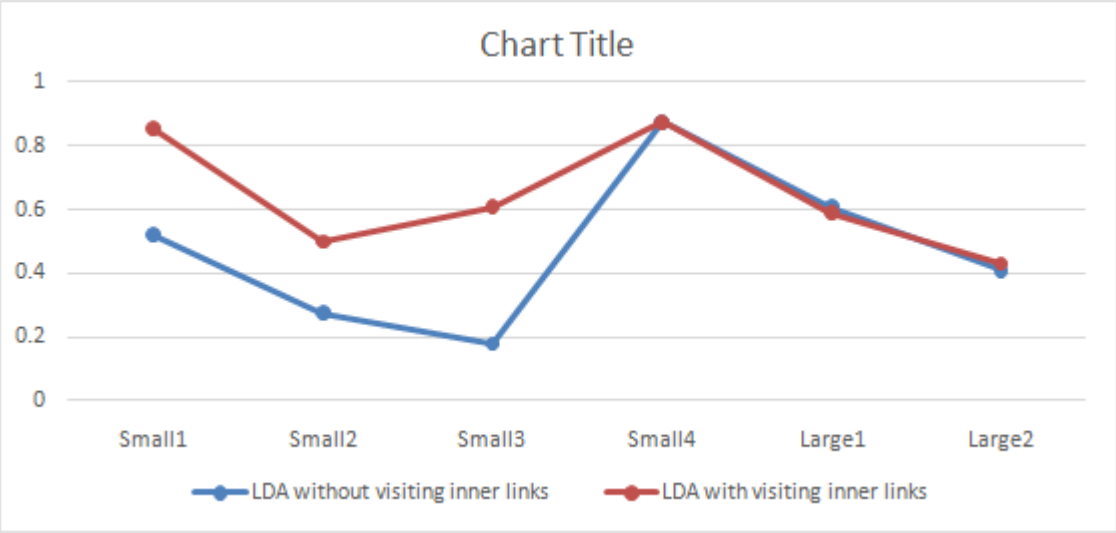


Fig 4: Performance evaluation for datasets

The above chart further strengthens the proof that the new method of visiting inner HTML links works better.

6. SUMMARY AND CONCLUSION

The requirement for an efficient and scalable clustering algorithm for deep web pages is fulfilled by our approach. The results show that sampling of data is useful for clustering heterogeneous deep Web sources. Our proposed Gibbs-LDA based technique is more efficient than the existing techniques.

It gives more accurate results when links in the web pages are parsed and their web page content are also taken into account. As LDA produce soft clusters it assigns probability to each document for each cluster. Hence, our tool is suitable for the scenario where the sources are sparsely distributed over the web.

In order to increase the weightage for the content in our base webpage over the one that we parse in our web page, we have increased the frequency of all the texts that appear in base webpage, which increases the space complexity. New innovative methods could be proposed in future which could possibly reduce space.

ACKNOWLEDGEMENTS

We would like to thank our Project guide, Dr. E. Sivasankar and Ms.Selvi Chandran, Research Scholar, CSE, NIT-Trichy. Their input, guidance and encouragement made this project possible and a success.

REFERENCES

[1] The Deep Web: Surfacing hidden value. Accessible at http://brightplanet.com, (2000).
[2] Madhavan, J., Cohen, S., Dong, X. L., Halevy, A. Y., Jeffery S. R.,Ko, D. and Yu, C (2007),"Web scale data integration", Conference on Innovative Data System Research (CIDR), 342–350.
[3] Tor: The Second-Generation Onion Router. Accessible at www.onion-router.net/Publications/tor-design.pdf
[4] Vincenzo Ciancaglini, Marco Balduzzi, Max Goncharov, and Robert McArdle, "Deepweb and Cybercrime: It's Not All About TOR", Trend Micro
[5] David M. Blei, Probabilistic topic models (2012), communications of the acm, Vol.55, No.4
[6] Estivill-Castro, Vladimir (2002), "Why so many clustering algorithms: a position paper", ACM SIGKDD Explorations Newsletter 4.1
[7] Blei, D., M., Ng, Y., A. and Jordan, M., I. (2003)," Latent Dirichlet Allocation", Journal of Machine Learning Research.
[8] Umara Noor, Ali Daud, Ayesha Manzoor (2013)," Latent Dirichlet Allocation based Semantic Clustering of Heterogeneous Deep Web Sources", In: Intelligent Networking and Collaborative Systems (INCoS), 132- 138.
[9] Unsupervised Learning. Accessible at http://mlg.eng.cam.ac.uk/zoubin/papers/ul.pdf, (2004).
[10] Thanaa M. Ghanem and Walid G. Aref. (2004), "Databases deepen the web". IEEE Computer, 37(1), 116–117.
[11] Dennis Fetterly, Mark Manasse, Marc Najork, and Janet Wiener (2004) "A large-scale study of the evolution of web pages." In Proceedings of the 12th International World Wide Web Conference, 669–678.
[12] Steve Lawrence and C. Lee Giles (1999), "Accessibility of information on the web." Nature, 400(6740), 107–109
[13] A Theoretical and Practical Implementation Tutorial on Topic Modeling and Gibbs Sampling. Accessible at http://u.cs.biu.ac.il/~89-680/darling-lda.pdf, (2011)
[14] The ElementTree XML API. Accessible at https://docs.python.org/2/library/xml.etree.elementtree.html
[15] The Bayesian Approach. Accessible at https://www.cs.cmu.edu/~scohen/psnlp-lecture6.pdf
[16] Google App Engine: Platform as a service. Accessible at https://cloud.google.com/appengine/docs
[17] Stop Words. Accessible at http://www.webconfs.com/stop-words.php
[18] The UIUC Web integration repository. Accessible at http://metaquerier.cs.uiuc.edu/repository
[19] Precision and recall .Accessible at en.wikipedia.org/wiki/Precision_and_recall.
[20] Clustering: Overview and K-means algorithm. Accessible at http://www.cs.princeton.edu/courses/archive/spr11/cos435/Notes/clustering_intro_kmeans_topost.pdf

INTENTIONAL BLANK

TURNOVER PREDICTION OF SHARES USING DATA MINING TECHNIQUES: A CASE STUDY

Shashaank D.S, Sruthi.V, Vijayalashimi M.L.S and
Shomona Garcia Jacob

Department of Computer Science and Engineering, SSNCE, Chennai, India.
shashaank.sivakumar@gmail.com
sruthivenkatesh1@gmail.com
vijayalakshimisethuraman@gmail.com
shomonagj@ssn.edu.in

ABSTRACT

Predicting the Total turnover of a company in the ever fluctuating Stock market has always proved to be a precarious situation and most certainly a difficult task at hand. Data mining is a well-known sphere of Computer Science that aims at extracting meaningful information from large databases. However, despite the existence of many algorithms for the purpose of predicting future trends, their efficiency is questionable as their predictions suffer from a high error rate. The objective of this paper is to investigate various existing classification algorithms to predict the turnover of different companies based on the Stock price. The authorized dataset for predicting the turnover was taken from www.bsc.com and included the stock market values of various companies over the past 10 years. The algorithms were investigated using the 'R' tool. The feature selection algorithm, Boruta, was run on this dataset to extract the important and influential features for classification. With these extracted features, the Total Turnover of the company was predicted using various algorithms like Random Forest, Decision Tree, SVM and Multinomial Regression. This prediction mechanism was implemented to predict the turnover of a company on an everyday basis and hence could help navigate through dubious stock markets trades. An accuracy rate of 95% was achieved by the above prediction process. Moreover, the importance of the stock market attributes was established as well.

KEYWORDS

Data mining, Feature selection, classification algorithms, Machine learning algorithms

1. INTRODUCTION

Prediction of stock market prices, its rise and fall of values has constantly proved to be a perilous task mainly due to the volatile nature of the market[1-3]. However data mining techniques and other computational intelligence techniques have been applied to achieve the same over the years. Some of the approaches undertaken included the use of decision tree algorithm, concepts of neural networks and Midas[4-6]. However through this paper, a comparative study was conducted to estimate and predict the turnover of companies that include Infosys, Sintex, HDFC

David C. Wyld et al. (Eds) : ACITY, DPPR, VLSI, WiMNET, AIAA, CNDC - 2015
pp. 29–38, 2015. © CS & IT-CSCP 2015 DOI : 10.5121/csit.2015.51303

and Apollo hospitals using various machine learning algorithms such as Random Forest, Decision Tree, Support Vector Machine and Multinomial Logistic Regression. In order to estimate the performance of the aforementioned machine learning algorithms in predicting the turnover, a confusion matrix was also constructed with respect to the dataset. Based on the predictions made by each of the algorithms with respect to the total turnover for a company (on an everyday basis), an accuracy rate was estimated for each of them from the number of true positives/negatives and false positives/negatives. A brief review of the state-of-the-art in predicting stock market share data is given below.

2. RELATED WORK

The objective of any nation at large is to enhance the lifestyle of common man and that is the driving force to undertake research to predict the market trends [7-9]. In the recent decade, much research has been done on neural networks to predict the stock market changes [10].

Matsui and Sato [12] proposed a new evaluation method to dissolve the over fitting problem in the Genetic Algorithm (GA) training. On comparing the conventional and the neighbourhood evaluation they found the new evaluation method to be better than the conventional one in terms of performance. Gupta, Aditya, and Dhingra [13] proposed a stock market prediction technique based on Hidden Markov Models. In that approach, the authors considered the fractional change in stock value and the intra-day high and low values of the stock to train the continuous Hidden Markov Model (HMM). Then this HMM is used to make a Maximum a Posteriori decision over all the possible stock values for the next day. The authors applied this approach on several stocks, and compared the performance to the existing methods. Lin, Guo, and Hu[14] proposed a SVM based stock market prediction system .This system selected a good feature subset, evaluated stock indicator and controlled over fitting on stock market tendency prediction. The authors tested this approach on Taiwan stock market datasets and found that the proposed system surpassed the conventional stock market prediction system in terms of performance.

3. PROPOSED STOCK TURNOVER PREDICTION FRAMEWORK

The stock turnover prediction framework proposed in this paper is portrayed in Figure 1. The basic methodology involved Data Collection, Pre-processing, Feature Selection and Classification, each of which is explained below.

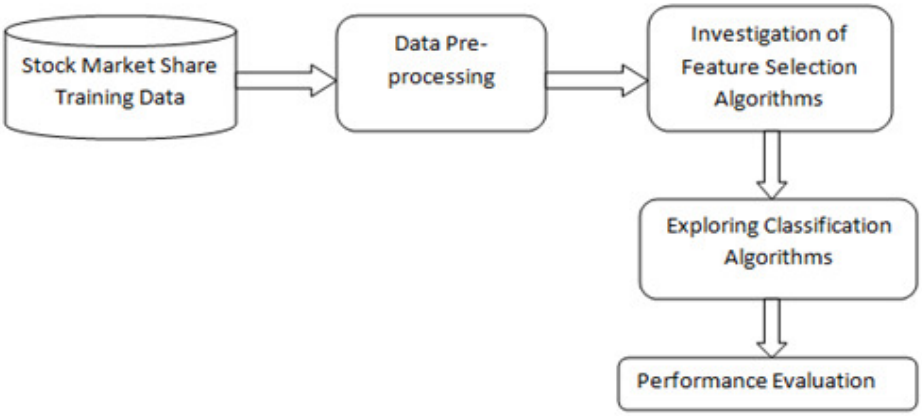


Figure 1: Stock Turnover Prediction framework

The dataset utilised for predicting the turnover was taken from www.bsc.com which included the stock market values of companies including Infosys, HDFC, Apollo Hospitals and Sintex, over the past 10 years.

3.1 Data Processing

Initially, all the records with missing values were removed from the dataset in order to improve the accuracy of the prediction. Then the data was further partitioned into two parts:

Training data (d.t): It is the data with which the machine is trained. Various classification algorithms are trained on this data. 60% of the data is taken as training data.

Validation data (d.v): It is the data which is used for the purpose of cross-validation. It is used to find the accuracy rate of each algorithm. The remaining 40% of the data is taken as validation data.

In order to apply the classification algorithms, the data was first sorted according to the turnover. Then the total turnover was discretised into:

A - 58,320 to 18,291,986

B – 18,296,597 to 37,731,606

C – 37,749,751 to 121,233,543

D – 121,245,870 to 300,360,881

E- 300,465,316 to 19,085,311,470

Also, the company features were converted into dummy variables (0's/1's) to help the prediction process easier.

The stock market data was characterised by attributes described in Table 1. The stock market starts at 9:15 in the morning and ends at 3:30 in the afternoon. The attributes described in Table 1 are recorded within this time frame.

Table 1. Stock Market Share Data – Attribute Description

S.NO	***ATTRIBUTE***	***DESCRIPTION***
1.	Open price	The first traded price during the day or in the morning.
2.	High price	The highest traded price during the day.
3.	Low price	The lowest price traded during the day.
4.	Close price	The last price traded during the day.
5.	WAP	Weighted average price during the day.
6.	No of shares	The total number of shares done during the day.
7.	No of trades	No of trades is the total no of transactions during the day.
8.	Deliverable quantity.	The quantity that can be delivered at the end of the day.
9.	Spread high low	Range of High price and low prices.
10.	Spread close open	Range of close and low prices.
11.	Company	The name of the company that handles the shares.
12.	Total turn over	Turnover is the total no of shares traded X Price of each share sold.
13.	Date	The date for which the above attributes are recorded.

Once the data was pre-processed, the important features to make an accurate prediction were identified by the process of feature selection.

3.2 Feature Selection

In order to estimate the possible influence of each of the above attributes on the predicted turnover, Boruta algorithm in R tool [15] was used. Boruta is a machine learning algorithm used to find relevant and important features for a given dataset i.e used to solve the minimal-optimal problem. The minimal – optimal problem is an often found situation today where most of the variables in a dataset are irrelevant to its classification. This problem gives rise to various disadvantages including over consumption of resources, slowdown of machine learning algorithms and most importantly, decrease in accuracy yielded by the same. Additionally, Boruta is a wrapper algorithm built around the Random Forest algorithm(implemented in the R package RandomForest) such that in every iteration the algorithm removes the irrelevant or less important features or attributes on the basis of the results rendered by a series of statistical tests.

The Boruta algorithm follows the following steps:

- The information system is expanded by adding duplicates of all attributes. These duplicates are known as shadow attributes.
- The added attributes are shuffled and the randomForest algorithm is run on the expanded information system. The resultant Z scores are noted.
- The Z score of every attribute is considered and the maximum Z score among all the shadow attributes (MZSA) is estimated. Further a hit value is assigned to every attribute that possesses a Z score better than MZSA.
- For each shadow attribute with undetermined importance perform a two-sided test of equality with the MZSA is conducted.
- All the attributes which have significantly lower importance than MZSA are considered to be 'unimportant' and permanently removed them from the information system.
- Similarly those attributes that having higher importance when considered alongside MZSA are considered to be important.
- All duplicates from the information system are removed.
- This procedure is repeated until the level of importance is assigned for all the attributes.

3.2 Sample code:

```
f.la <- Total_Turnover ~.
at.select <- Boruta (formula = f.la, data = d.t)
at.select$finalDecision
```

And, the graph obtained is as given in figure 2

3.3 Classification

Classification [16-17] is the process of finding a set of models that describe and distinguish data classes. This is done to achieve the goal of being able to use the model to predict the class whose label is unknown. The classification phase involved the execution of the classification algorithms to identify the best performing algorithm. The classification accuracy obtained by percentage

split as discussed in the data pre-processing phase, was calculated and a comparison was drawn among the classifiers. The algorithms that yielded the highest accuracy is described below.

Random Forest

In random forest [18-19] a randomly selected set of attributes is used to split each node. Every node is split using the best split among a subset of predictors that are deliberately chosen randomly at the node. This is in contrast to the methodology followed in standard tress in which each node is split using the best split among all attributes available in the dataset. Further new values are predicted by aggregating and collating the predictions of the various decision trees constructed.

Random forest represents an ensemble model / algorithm as it derives its final prediction from multiple individual models. These individual models could be of similar or different type. However, in the case of Random Forest, the individual models are of the same type – decision trees.

Sample code:

```
f.la <- Total_Turnover ~.
dt.fit <- randomForest(formula = f.la,data = d.t)
dt.fit.v <- predict (object = dt.fit,newdata = d.v ,type = 'class')
```

The Random Forest algorithm yielded 95.08% accuracy with all the 12 features, the results of which are discussed in the ensuing section.

The comparative performances of the feature selection and classification algorithms are discussed below.

4. RESULT ANALYSIS

The results analysis is discussed in two sections. The former section elaborates on the feature selection process while the latter section makes a detailed analysis on the performance of the classification algorithms.

4.1 Performance Analysis of Feature Selection

As discussed before, the Boruta package was utilized for performing feature selection on the pre-processed training data. The graphical representation of the importance of the features and their role in enhancing the classification accuracy is portrayed in Figure 2.

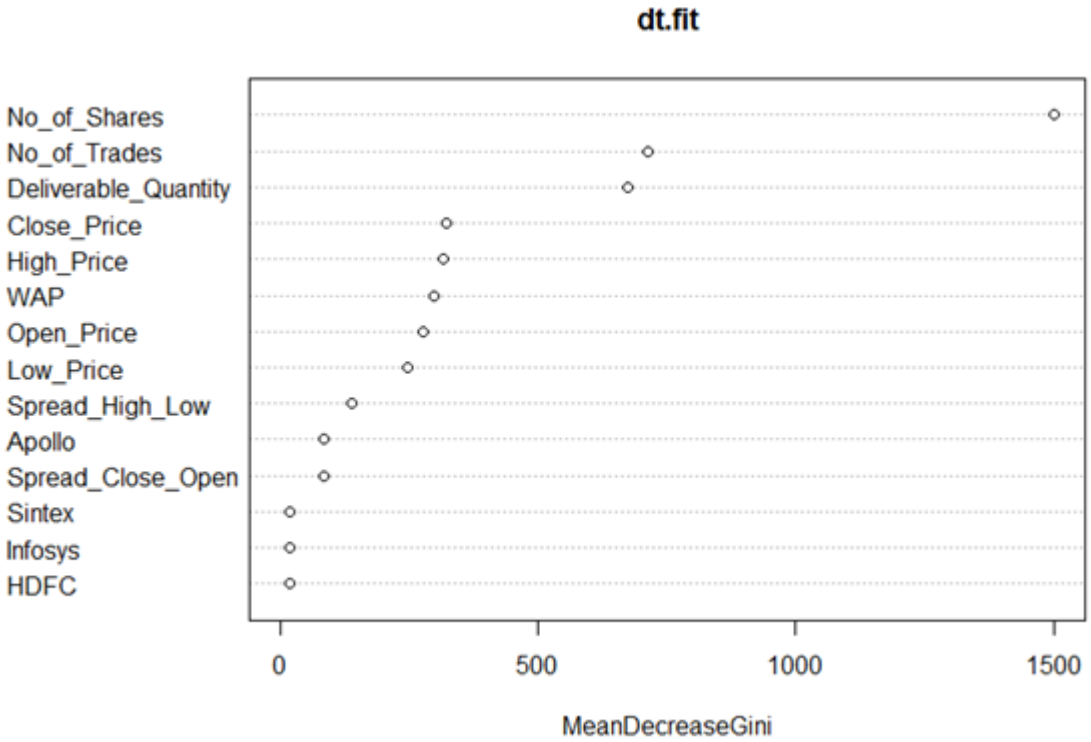


Figure 2: Attribute Importance in Share Turnover Prediction – Boruta Package

Once the important features were identified, the next phase involved predicting the turnover from the features in order to estimate the probable combination of attributes that yield a high turnover.

4.2 Performance Analysis of Classification Algorithms

Each of the classification algorithms were first trained using the training data which contained 60% of the dataset. Then the remaining 40% was used for the purpose of cross-validation. From the prediction produced by each of the algorithms, the confusion matrix was obtained. Further the accuracy rate was determined by the formula:

$$\text{Accuracy rate} = \frac{\text{No. of correctly classified observations}}{\text{Total No. of observations}} \times 100$$

Table 2. Comparative Performance of Classification Algorithms

S.No	Classification Algorithms	Accuracy (%)
1	Random Forest	95.08
2	Decision tree – PARTY	89.5
3	Decision tree- Rpart	82.3
4	SVM	75.9
5	MLR	73.55

The graphical representation of the total turnover prediction of the companies is given in Figure 3.

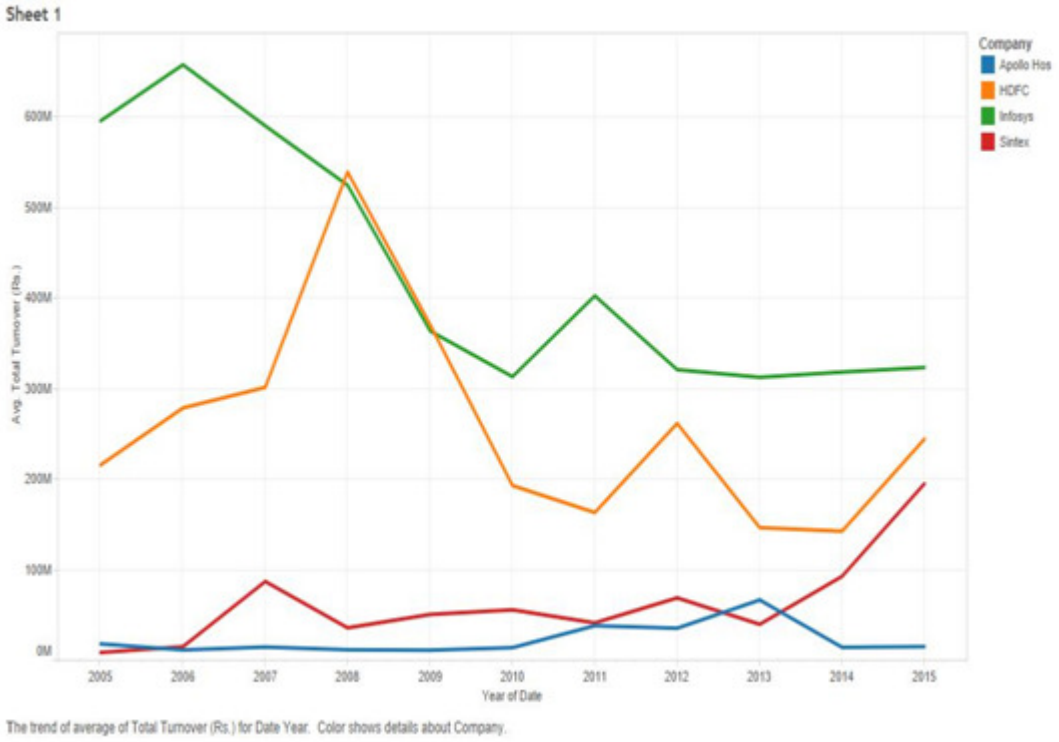


Figure 3 .Overall Prediction of Turnover for the Companies

The line graph was plotted between the Average turnover for each year and the corresponding year itself. The details of the graph shown in Figure 3 are as follows:

- Apollo hospital: Initially increased from in 18,020,386 in 2005 to 66,527,438 in 2013 and then decreased to 14,995,685 in 2015.
- Infosys: Initially decreased from 595,911,899 in 2005 to 313,287,090 in 2010, increased to 402,652,597 in 2011 and further decreased to 323,405,496 in 2015.
- Sintex: Increased with a few variations from 8,495,478 in 2005 to 194,974,362 in 2015.
- HDFC: Initially increased from 251,751,341 in 2005 to 539,536,979 in 2008 and then decreased to 244,309,549 in 2015.

The bar graph was plotted between 9831 opening prices grouped by total turnover with the discretized value of the total turnover. The graph clearly stated that there total turnover did not increase linearly with respect to the sum of the No of shares. For example, the No of shares for D was 529,957,607 whereas, for E it is 617,959,679.

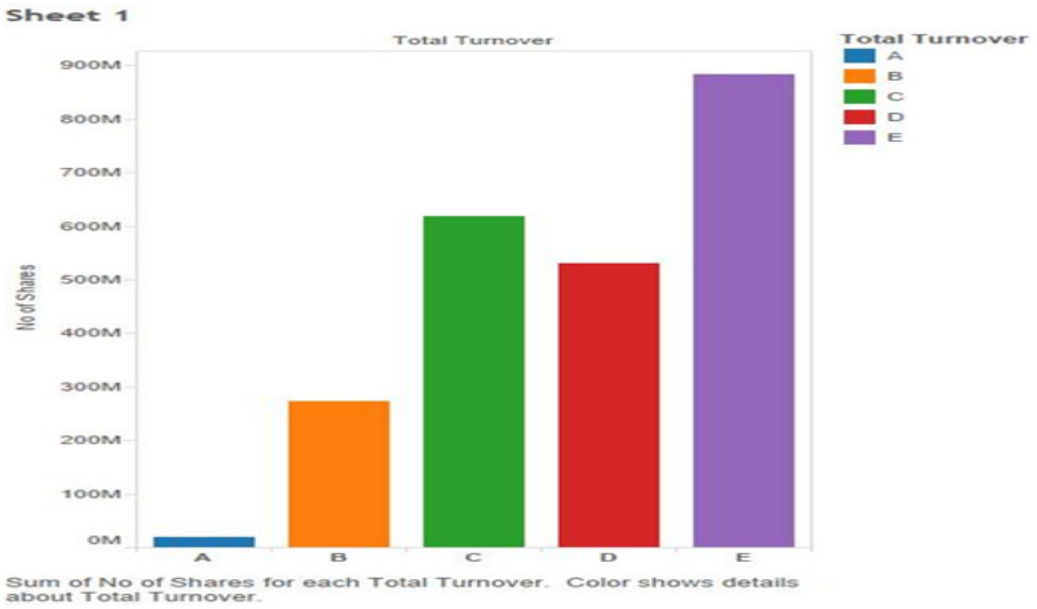


Figure 4. Variation of Total_Turnover with Respect to No_Of_Shares

It is evident from the result analysis that almost all the features are required to accurately predict the total turnover of a company. Moreover the Random Forest algorithm has proved to accurately predict the turnover for the real-time share data of different companies which gives the lead to investigate many other boosting and ensemble techniques to enhance the prediction accuracy.

5. CONCLUSION

Application of data mining techniques to predict turnover based on stock market share data is an emerging area of research and is expected to be instrumental in moulding the country's economy by predicting possible investment trends to increase turnover. In view of this, an efficient way of implementing the Random Forest algorithm is proposed in order to mitigate the risks involved in predicting the turnover of a company. It was also identified that all features involved in the stock marker share data were essential for prediction. An accuracy rate of 95% was achieved in the prediction process. This accuracy rate was much higher than those obtained before. Hence we believe that further research using computational methodologies to predict turnover on a daily basis based on share market data will reveal better and more interesting patters for investments.

REFERENCES

[1] Abhishek Gupta, Dr. Samidha , D. Sharma - "Clustering-Classification Based Prediction of Stock Market Future Prediction"- (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (3) , 2014, 2806-2809.

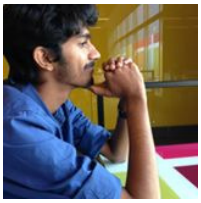
[2] Dharamveer , Beerendra , Jitendra Kumar –" Efficient Prediction of Close Value using Genetic algorithm based horizontal partition decision tree in Stock Market " Volume 2, Issue 1, January 2014 International Journal of Advance Research in Computer Science and Management Studies Research Paper Available online at: www.ijarcsms.com.

[3] Kannan, K. Senthamarai, et al. "Financial stock market forecast using data mining techniques." Proceedings of the International Multiconference of Engineers and computer scientists. Vol. 1. 2010.

[4] Md. Al Mehedi Hasan, Mohammed Nasser, Biprodip Pal, Shamim Ahmad – "Support Vector Machine and Random Forest Modeling for Intrusion Detection System (IDS)"- Journal of Intelligent Learning Systems and Applications, Vol.6 No.1(2014), Article ID:42869,8 pages
[5] Shen, S., Jiang, H., & Zhang, T. (2012). Stock market forecasting using machine learning algorithms.
[6] Bayaga, Anass. "Multinomial logistic regression: usage and application in risk analysis." Journal of applied quantitative methods 5.2 (2010): 288-297.
[7] Shah, Vatsal H. "Machine learning techniques for stock prediction." Foundations of Machine Learningl Spring (2007).
[8] Al-Radaideh, Qasem A., Aa Assaf, And Eman Alnagi. "Predicting Stock Prices Using Data Mining Techniques." The International Arab Conference on Information Technology (ACIT'2013). 2013.
[9] Wang, Jar-Long, and Shu-Hui Chan. "Stock market trading rule Discovery using two-layer bias decision tree." Expert Systems withApplications 30.4, 605-611, 2006.
[10] Wu, Muh-Cherng, Sheng-Yu Lin, and Chia-Hsin Lin. "An effectiveApplication of decision tree to stock trading." Expert Systems withApplications 31.2: 270-274, 2006.
[11] Mahdi Pakdaman Naeini, Hamidreza Taremian, Homa Baradaran Hashemi "Stock Market Value Prediction Using Neural Networks", International Conference on Computer Information Systems and Industrial Management Applications (CISIM), pp. 132-136, 2010.
[12] Matsui, Kazuhiro, and Haruo Sato. "Neighbourhood evaluation in acquiring stock trading strategy using genetic algorithms." Soft Computing and Pattern Recognition (SoCPaR), 2010 International Conference of. IEEE, 2010.
[13] Gupta, Aditya, and Bhuwan Dhingra. "Stock market prediction using hidden markov models." Engineering and Systems (SCES) Students Conference on, IEEE, 2012.
[14] Lin, Yuling, Haixiang Guo, and Jinglu Hu,"An SVM-based approach for stock market trend prediction." Neural Networks (IJCNN), the 2013 International Joint Conference on. IEEE, 2013.
[15] Sanjana Sahayaraj, Shomona Gracia Jacob, Data Mining to Help Aphasic Quadriplegic and Coma Patients, International Journal of Science and Research (IJSR), Vol. 3(9), pp.121-125, 2014.
[16] Jacob SG, Ramani RG, Prediction of Rescue Mutants to predict Functional Activity of Tumor Protein TP53 through Data Mining Techniques", Journal of Scientific and Industrial Research, Vol.74, pp.135-140, 2015.
[17] Zhao, Yanchang. R and data mining: Examples and case studies. Academic Press, 2012.
[18] Andy Liaw and Matthew Wiener. "Classification and Regression by random forest"
[19] Nazeeh Ghatasheh. "Business Analytics using Random Forest Trees for Credit RiskPrediction: A Comparison Study".

AUTHORS

Shashaank D.S is currently pursuing B.E computer Science and Engineering in SSN College of Engineering Chennai, India. He is doing research in the field of machine learning and is interested in speech processing.



Sruthi.V is currently pursuing B.E computer Science and Engineering in SSN College of Engineering Chennai, India. She is doing research in the field of machine learning.

Vijayalakshimi M.L.S is currently pursuing B.E computer Science and Engineering in SSN College of Engineering Chennai, India. She is doing research in the field of machine learning.

Dr. Shomona Gracia Jacob is Associate Professor, Department of CSE, SSN College of Engineering, Chennai, India. She completed Ph.D at Anna University in the area of Biological and Clinical Data Mining. She has more than 30 publications in International Conferences and Journals to her credit. Her areas of interest include Data Mining, Bioinformatics, Machine Learning, and Artificial Intelligence. She has reviewed many research articles on invitation from highly reputed refereed journals. She is currently guiding under-graduate and post-graduate projects in the field of data mining and intelligent systems.

WEB MINING BASED FRAMEWORK FOR ONTOLOGY LEARNING

C.Ramesh[1], K.V.Chalapati Rao[1], A.Govardhan[2]

[1]Department of Computer Science and Engineering, CVR College of Engineering,
Ibrahimpatnam, R.R.District, Telangana, India.
hmcr.ramesh@gmail.com, chalapatiraokv@gmail.com
[2]School of IT, JNT University Hyderabad, Hyderabad,
Telangana, India.
govardhan_cse@yahoo.co.in

ABSTRACT

Today, the notion of Semantic Web has emerged as a prominent solution to the problem of organizing the immense information provided by World Wide Web, and its focus on supporting a better co-operation between humans and machines is noteworthy. Ontology forms the major component of Semantic Web in its realization. However, manual method of ontology construction is time-consuming, costly, error-prone and inflexible to change and in addition, it requires a complete participation of knowledge engineer or domain expert. To address this issue, researchers hoped that a semi-automatic or automatic process would result in faster and better ontology construction and enrichment. Ontology learning has become recently a major area of research, whose goal is to facilitate construction of ontologies, which reduces the effort in developing ontology for a new domain. However, there are few research studies that attempt to construct ontology from semi-structured Web pages. In this paper, we present a complete framework for ontology learning that facilitates the semi-automation of constructing and enriching web site ontology from semi structured Web pages. The proposed framework employs Web Content Mining and Web Usage mining in extracting conceptual relationship from Web. The main idea behind this concept was to incorporate the web author's ideas as well as web users' intentions in the ontology development and its evolution.

KEYWORDS

Ontology Learning, Web Mining, Web Content Mining, Web Usage Mining, Ontology Evaluation

1. INTRODUCTION

World Wide Web, since its conceptual inception, has contributed greatly for the knowledge era, in which we are living today. As conceptualized by Sir Tim Berners-Lee, the introduction of World Wide Web (WWW) has given rise to enormous amount of information that can be accessed in digital form and most of these data are in the form of documents. The exponential growth of these documents has raised many challenges. Considering the structure of these

David C. Wyld et al. (Eds) : ACITY, DPPR, VLSI, WiMNET, AIAA, CNDC - 2015
pp. 39–51, 2015. © CS & IT-CSCP 2015 DOI : 10.5121/csit.2015.51304

documents, we find that they are not descriptive enough to express themselves, overloaded with information and distributed all over the Web. Therefore, it has become a difficult task for the Web Users to search and retrieve the relevant information needed for them.

Semantic Web, as envisioned by Sir Tim Berners-Lee, addresses this problem by giving information a well-defined meaning, better enabling computers and people to work in co-operation. Semantic Web is implemented using W3C recommended Semantic Web Technologies and Standards and expresses the Web data in a machine-understandable and machine processable form, thereby supporting information exchange and sharing between applications. Ontologies play a significant role in building Semantic Web and provide a platform for promoting Semantic interoperability on the Web. However, constructing ontology's for the many and varied domains on the Web is a time-consuming process and their construction is a bottleneck to the wider deployment and use of Semantic Information on the Web. Since manual construction of ontology is costly, time-consuming, error-prone and inflexible to change, it is hoped that an automated or semi-automated process will result in better ontology construction and create ontologies that better match a specific application [1].

There have been several research attempts to automate ontology construction and update process by exploiting the content of Web pages. Most of the Web documents that exist today are in semi-structured format. However, there are few references to research attempts that focus on these semi-structured data on Web [2] [3] [4]. Further most of these research attempts use text mining and Natural Language Processing techniques to extract the semantics from Web documents, neglecting the embedded information in the semi-structured nature. Also most of the current approaches are dealing with some specific tasks or a part of the ontology learning process rather than providing complete support to users. There are few research attempts that use Web mining techniques such as Web Content Mining and Web Usage Mining in ontology development.

The benefits of analyzing the usage behavior analysis have been the driving forces for continuous research in the realm of Web Usage Mining, which aims at discovering navigational patterns from the logs of HTTP requests for Web resources [5]. Further Web Content Mining aims to extracts/mine useful information or knowledge from Web page contents. The benefits offered by these two techniques in Web Mining applications are noteworthy.

In this paper, we present a framework for Ontology Learning from Semi structured Web pages using the combined techniques of Web Mining namely, Web Content Mining and Web Usage Mining. We have employed the Web Content Mining to extracts the concepts and further discover the Conceptual relationships from Web pages. We applied the text mining techniques and extended Apriori Algorithm, which is most widely used for frequent mining, for extracting the concepts. The Semantics extracted from Web Usage Mining process, helps in refining the conceptual relationships extracted from Web Content Mining. Further the refined conceptual relationships are also used in enriching the Web site Ontology. Ontology Pruning and Ontology evaluation are other stages of Ontology Learning process.

The remainder of this paper is organized as follows. In section II, we present a survey of current research efforts on Ontology Learning and Web Mining Methods. In section III, we present our Ontology Learning framework and its main architectural components. In section IV, implementation and experimental results are discussed. In section V, enriched Ontology is evaluated. Finally, in conclusion, some plans for future work are presented.

2. RELATED WORK

"Ontology is an explicit, formal specification of a shared conceptualization of domain of interest [6], where formal implies that the ontology should be machine readable and the domain can be conceptual thing that is shared by a group or community". During the last decade, several research attempts on ontology learning and systems have been proposed. These research efforts tried to build ontology in either of two ways. One way is using ontology development tools [8] like protégé and Onto-Edit. Knowledge engineers and Domain experts use these tools to build the ontology. Another one is semi-automatic way of constructing the ontology by learning it from different information sources [9] [10] with little human intervention.

Ontology learning refers to a process of applying various knowledge discovery techniques in constructing ontology by extracting concepts and relations using different input sources. It aims at building ontologies semi automatically or automatically from a given text corpus with a limited human exert. Ontology learning can also be defined as a set of methods and techniques used for building ontology from scratch, enriching or adapting an existing ontology in a semi-automatic fashion using several sources [9]. Ontology learning has recently been studied as an effective approach to facilitate the semi- automatic development of ontologies. Ontology learning use techniques and methods from diverse spectrum of fields such as machine learning, knowledge acquisition, natural language processing, information retrieval, artificial intelligence, reasoning and database management systems[11][9].

Manual construction of ontologies is costly, time-consuming, error-prone and inflexible to change. Ontology learning systems can be categorized according to the type of data from which they are learned. Unstructured, fully structured and Semi-structured types of data especially form the input sources to ontology learning systems. In literature, there are several research attempts, focusing on constructing ontology for semi-structured Web Pages using various techniques. Research attempts that focus on unstructured Web pages [12][13][14][1] with free text, mostly use Natural Language Processing techniques and simple text mining in the ontology development. The research attempts that focus on fully structured Web Pages, such as Wikipedia, move beyond simple text mining and take into account the structure of the Web pages [15][16]. However, there are only few research efforts that focus on extracting Semantics from semi-structured Web pages.

The work presented in [3] was the first attempt to discuss the synergy between Semantic Web and Web Mining. They discussed the role of Web Mining techniques in facilitating ontology development. They claimed that the synergy between Semantic Web and Web Mining will give rise to a form of close loop learning, by allowing the usage of Web Mining to extract Semantics and building the Semantic Web and then using the Semantic structures in increasing the performance of Web Mining results. The work presented in [4] draw attention of researchers to use the mark up tags of HTML pages to be used in Web Content Mining to facilitate Ontology development. Descriptions of various techniques provided by Web Usage Mining in improving site Semantics and supporting the users in their navigation is well presented in [2].

A framework for Web Usage driven adaptation of the Semantic Web is well presented in work [17]. The adaptation process employed in the framework, exploited the Web access logs of the users, together with the semantic aspect of the Web to facilitate the Web browsing. Based on the usage of Web, they performed evolution of Web site ontology and topology. However in their

work, mining the content of the Web pages was not considered to full extent in extracting concepts needed for facilitating the ontology development.

In another approach [18], similar to our work, has presented a framework that combines Web Content Mining with Web Usage Mining to extract conceptual relationships hidden in semi-structured Web pages and used in ontology development. The main idea behind this concept was to incorporate the Web author's ideas as well as Web Users' intentions on Web site in the ontology development. The above research attempts to use Natural Language Processing and Association rule mining to extract the conceptual relationships. However, a complete ontology learning process was not presented and much focus was given only to ontology creation.

A semi-automatic method for domain terminology extraction from e-learning resources and their organization as ontology is well described in [19]. However, the work is limited only to e-learning domain and used mostly the Natural Language Processing techniques. Few research works that try to use the semi structured nature of the Web pages in ontology development have become specific to special type of Web sites such as template driven Web sites [20].

Research work [21] made use of the structure of phrases appearing in the HTML documents' headings, in identifying new concepts and taxonomical relationships. However, in most of the current research works, plain text is extracted from the semi structured Web pages as part of preprocessing phase and simple text mining techniques are applied on the extracted free text to construct ontology. Here the ontology development process has not considered the users' intentions and aspirations on Web site.

3. ARCHITECTURE

The main aim of the paper is to investigate and develop a framework that enables ontology learning by partially automating the process of extracting conceptual relationships from semi structured Web pages using Web Mining techniques. In this section, we present the overall architecture of our Ontology Learning framework. Figure.1 shows the architecture of our proposed Ontology Learning framework, consisting of four stages.

They are :

i. Mining the Web Page Contents to extract the Concepts and Conceptual relationships,

ii. Mining the Web Usage data to extract hidden Conceptual knowledge and refine the Conceptual relationships discovered in step one,

iii. Ontology construction and

iv. refining the Web site ontology. The input for the proposed Ontology Learning framework consists of site Web pages, server's access logs, the initial domain ontology of the Web site.

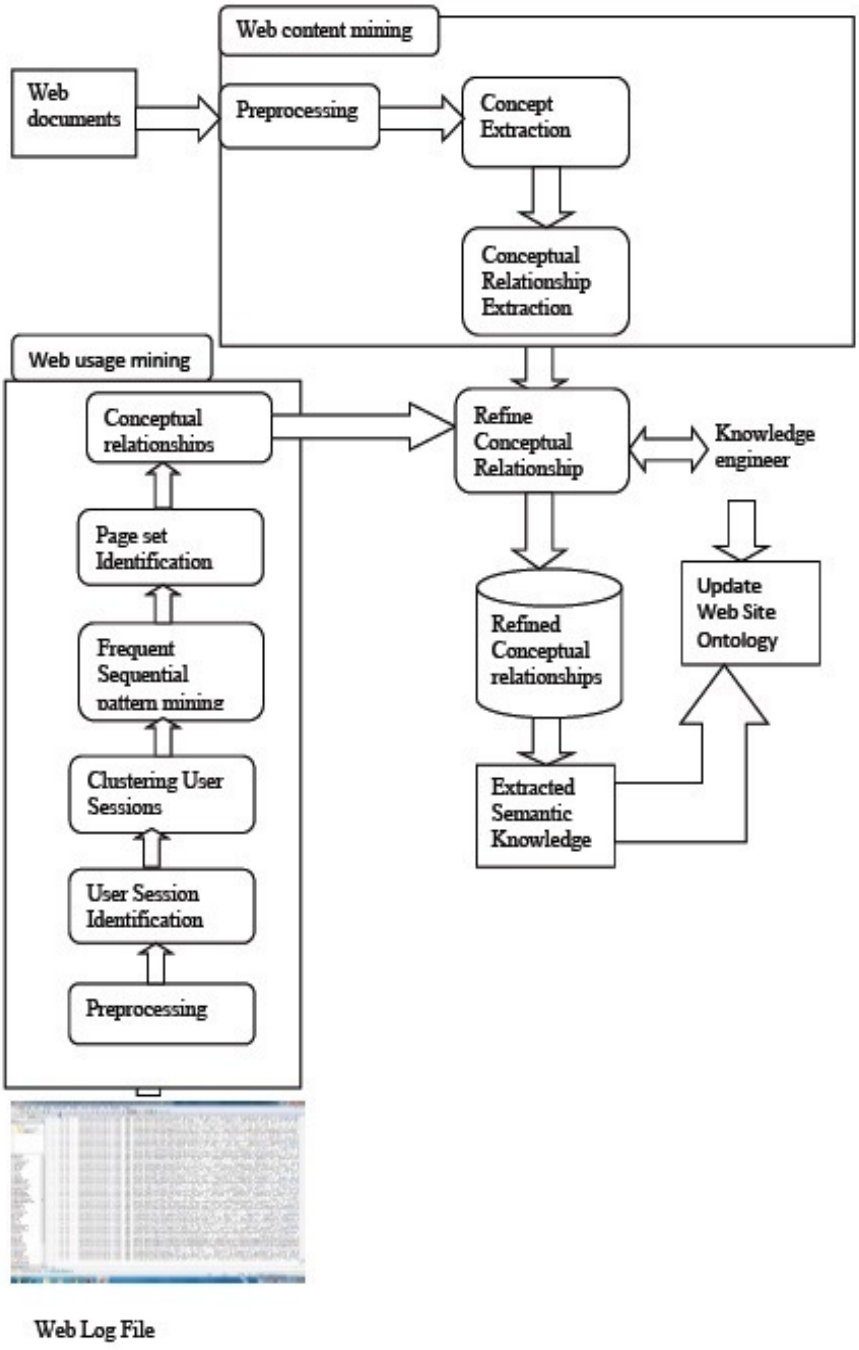


Figure.1. Architecture of Ontology Learning Framework

3.1 Mining the Web Page Content to Extract the Knowledge

Usually the Web page contents are organized from Web designer / Web author perspective. Mining the web page contents can reveal the conceptual relationships as seen by Web author. However extracting information directly from Web pages is a difficult task since most of the Web pages do not confirm to HTML syntax. The ill-formed HTML pages need to be preprocessed and parsed before applying the concept extraction process. A Java based SAX parser is used to parse the Web pages and convert them into a formal structure. The Web pages are annotated with parts of speech tags. Weighted Frequency of the concepts is determined by considering the frequency of the concepts as well as the frequency of HTML markers containing those concepts. Here different HTML tags are given different weights to match their importance.

The concepts that have a weighted frequency higher than externally specified threshold values are considered as most significant set of concepts of that Web page. One or more keywords of sentence in Web page may define a concept. An extended Apriori algorithm [22] was used to determine the significant concept sets, while pruning the word sequences from the candidate word sequences that have no probability of selecting as a concept. Concept sets are generated using the above process iteratively.

Associations between the concepts are identified as the concepts that together occur frequently. These associations between the concepts hint the existence of conceptual relationships. The identified associations mostly represent the conceptual relationships that exist in Web author or Web designer mind. These extracted set of conceptual relationships are presented to the ontology developer for further refinement where he/she can include any new conceptual relationships or remove irrelevant ones from the extracted conceptual relationships and refine the existing Web site ontology. Association rule discovery techniques are used in extracting the frequent concept sets. The most widely used, most popular CloSpan algorithm [23] was employed for extracting frequent concept sets. Conceptual relationships are determined from the generated association rules.

3.2 Mining the Usage Patterns to Extract Conceptual Relationships

Web Usage Mining refers to the process of extracting users' navigational patterns by applying data mining techniques on Web access log files. Users' Web Browsing activity is heavily dependent on his needs, knowledge and interests. Users' view on Web pages could be different from Web site author views. Mining the Usage patterns could reveal the conceptual relationships that reside in the web pages as per Web users' perspective. Web Usage patterns are used in applications like Web Personalization, Web caching, Web perfecting, Web site restructuring and intelligent online advertisements [24].

Web Users browsing preferences could be learned and adopted in the Web adaptation process to suit the users' needs. The Proposed framework uses Web Usage Mining to extract conceptual relationships that could be learnt about the Web pages according to the discovered usage patterns. The extracted Semantics is used in the conceptual relationships' refinement stage along with the conceptual relationships extracted by mining the Web content of the Web pages. Web Usage Mining alone cannot be used in extracting the Semantic Knowledge from Web access logs as the users' navigational patterns would be insufficient in case of dynamic Web sites where the content of the Web pages changes frequently.

Web Usage Mining process mainly includes steps like preprocessing the web log files, User Session Identification, discovery of frequent Sequential Patterns, Analysis of the Usage patterns and uses the discovered patterns in various applications.

3.2.1 Preprocessing the Access Log Files

The irrelevant information that exists in the raw Web access log files has to be removed before applying the Mining techniques. Here various preprocessing tasks are done. Removing duplicate records and irrelevant requests such as request with response status code greater than 200 and removing records related to image requests are done as part of the preprocessing task.

3.2.2 User Session Identification

After preprocessing phase, user sessions are identified. We used a heuristic measure in performing sessionization. An idle time of 30 minutes is considered in constructing user sessions. The identified user sessions are then mapped into multidimensional vector space model of URL references (bit vector).We represented each Web page visited as '1' and each Web page not visited as '0' while mapping the user sessions into a vector space model. Table.1 illustrates the user sessions mapping into multi dimensional vector space model.

Table 1. Example of User Sessions Mapping to Multi dimensional Vector space

User Session	Web Transaction Set
S1 = < p1,p2,p4,p5>	W1 = <1,1,0,1,1>
S2 = <p2,p3,p5>	W2 = <0,1,1,0,1>
S3 = <p1,p3,p5>	W3 = <1,0,1,0,1>

The constructed vector space is clustered using K-means clustering algorithms. Each cluster represents a group of Web transactions that are similar based on the co-occurrences of the URL references. The results are presented to the ontology developer who specifies the number of clusters to be generated. Sequential association rule mining is applied on the cluster sessions. Table.2 shows an example of the cluster details.

Table 2. Example of a Cluster details

Property	Value
1	{(1,0,0,0)(1,1,0,0)}
2	{(1,1,1,1) (0,0,0,1)}
3	{(1,0,0,1)}

3.2.3 Sequential Frequent Pattern Mining

Page sets are extracted using association rules. Based on the extracted page sets, conceptual relationships are identified and then presented to the ontology developer for suggestions. The ontology developer extracts useful conceptual relationships, which refine the Web site ontology.

Then the extracted information has to be converted into machine understandable format. Owl is used to represent the Semantic information.

4. EXPERIMENT AND RESULTS

Experiments are conducted on an anonymous University Web site. We have collected the Web access log file over a period of one month from University Server. For performing experiments, we used domain ontology of the same anonymous University Web site. Figure.2 shows the snapshot of initial domain ontology of the University Web site. The size of raw Web log file collected was nearly around 25540 page views. After preprocessing the log, the Number of page views, are reduced to 6892. The Number of User Sessions obtained were 600 on an average basis of 80 sessions per day.

The ontology was edited and visualized using tool Protégé'4.3 [25]. OWL language was used in representing the updated ontology. After preprocessing task, User sessions were identified. K-means Clustering algorithm is employed to generate clusters over generated User Sessions. CloSpan algorithm was implemented on the usage clusters to generate frequent sequential concept sets.

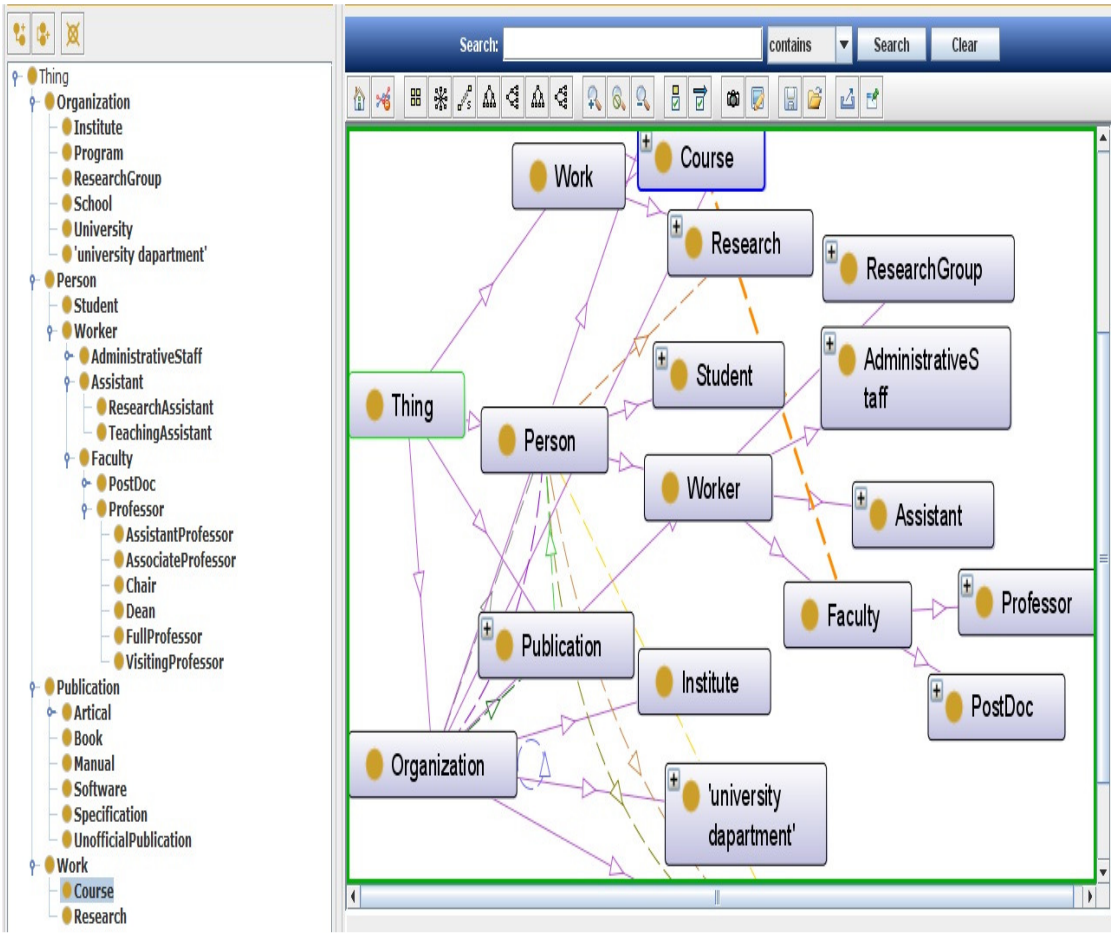


Figure 2. Domain Ontology of an Anonymous University Web site

We report in this section, some of the Sequential Association rules extracted in the Web Usage Mining process.

Pattern 1:
/person/student.html -> /person/worker/faculty/professor.html
-> /person/worker/teachingasst.html

Support: 0.02537

This behavior has a support of 2.53%. It corresponds to approximately 147 users of the web site. These users are likely to be interested in looking for teaching assistant. The support count of the above Association rules, hints that the existence of strong relationship between 'student' concept and 'teachingassistant' concept.

Pattern 2:

/person/worker/faculty/lecturer.html-> /person/worker/assistant/researchasst.html
-> /organization/researchgroup/course.html

Support: 0.02245

This behavior has a support of 2.20%. It corresponds to approximately 120 users of the web site. These users are likely to be interested in browsing research pages after viewing lecturer pages. It hints the existence of association between lecturer and research. Figure.3 shows the snippet code of the updated ontology in OWL representation.

```
3   xmlns = "http://www.jntuh.ac.in/~cse/2015/0401/univ-bench.owl#"
4   xml:base = "http://www.jntuh.ac.in/~cse/2015/0401/univ-bench.owl"
5   xmlns:rdf = "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
6   xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
7   xmlns:owl="http://www.w3.org/2002/07/owl#">
8  <owl:Ontology rdf:about="">
9   <rdfs:comment>An university ontology for benchmark tests</rdfs:comment>
10  <rdfs:label>Univ-bench Ontology</rdfs:label>
11  <owl:versionInfo>univ-bench-ontology-owl, ver April 1, 2015</owl:versionInfo>
12 </owl:Ontology>
13 <owl:Class rdf:ID="Employee">
14  <rdfs:label>Employee</rdfs:label>
15  <owl:intersectionOf rdf:parseType="Collection">
16  <owl:Class rdf:about="#Person" />
17  <owl:Restriction>
18  <owl:onProperty rdf:resource="#worksFor" />
19  <owl:someValuesFrom>
univ-bench1.owl
```

Figure 3. Snapshot of an updated ontology in OWL representation

Reorganization of the concepts hierarchy was performed. For instance, the sub-concept "university research assistant" was previously not under the concept "Student ". However, the high frequency with which these two concepts were occurring together hints the existence of conceptual relationship between them and reflects the interest of the users. The result ontology is represented in OWL language. The resulting ontology after the application of changes as suggested by the system is shown in Figure 4.

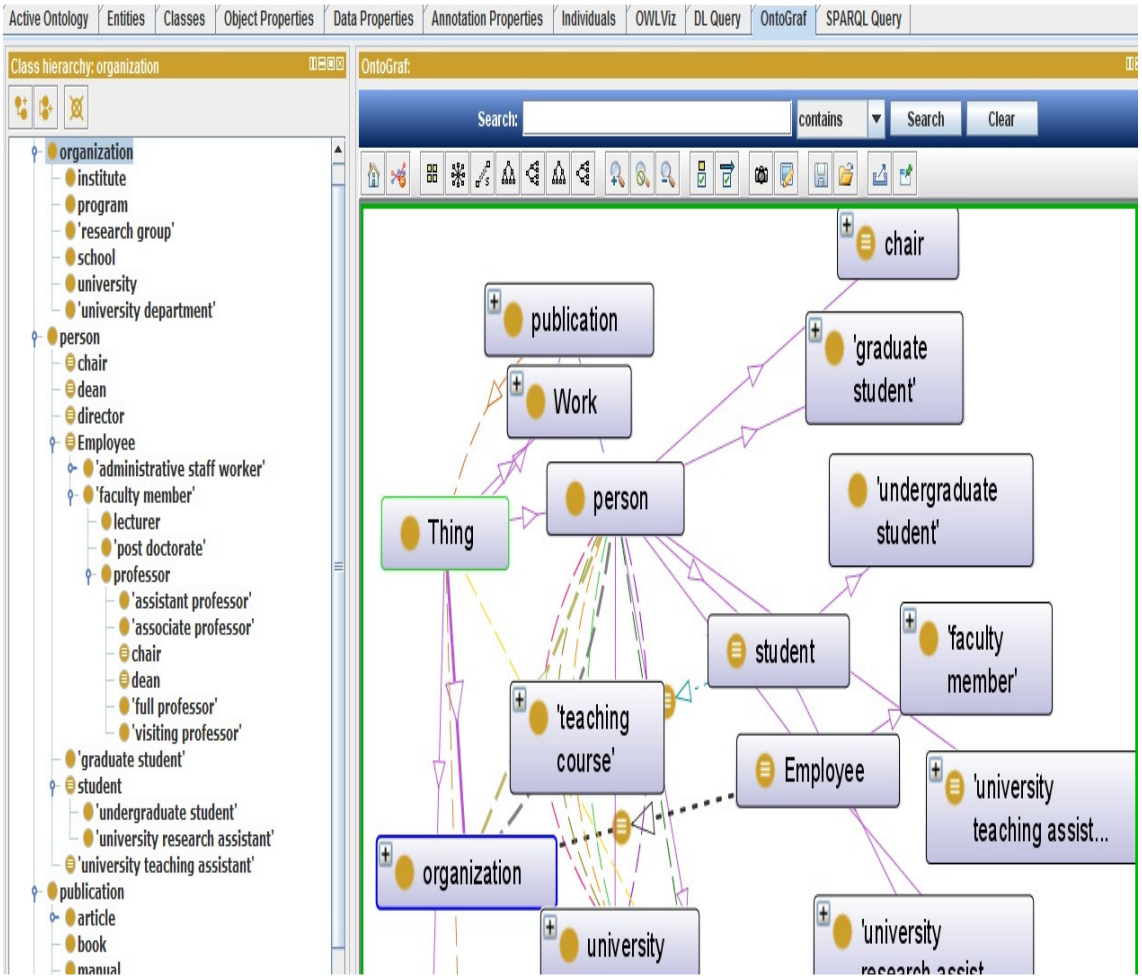


Figure 4. Updated University Ontology after incorporating the extracted relevant conceptual relationships proposed by the framework

4.1 Ontology Evaluation

Ontology evaluation is critically important if the ontologies are to be widely adopted in the Semantic Web and other Semantic aware applications. Evaluation of ontology refers to a process of assessing whether a given ontology represents the conceptual relationships in a given domain or selecting the best ontology among the candidate ontologies. One can assess the quality of the constructed ontology. In this paper we have assessed the quality of semi-automatically developed ontology. A comprehensive survey of existing ontology evaluation techniques is presented in research work [26]. Four types of evaluation approaches were discussed in the above work.

- Gold Standard Evaluation: Comparing the ontology based on a gold standard.
- Application based evaluation: based on the results obtained by using the ontology in an application.
- Data driven evaluation: based on the comparisons done on the source data.
- User evaluation: based on the evaluation done by humans.

Based on the nature of our research, user evaluation is well suited to assess the quality of the extracted Semantics because it is the user (knowledge engineer) who finally decides whether to add a particular conceptual relationship to web site ontology or not, and he/she is the best person to judge the quality of the extracted semantics. The report of the conceptual relationships is presented to ontology developer who decides for updating the ontology. He assesses the quality of generated ontology or enriched ontology.

5. CONCLUSIONS AND FUTURE WORK

Ontology is regarded as a backbone for Semantic Web. Manual acquisition of ontologies is tedious and cumbersome task. Constructing ontology for new domain is time-consuming and costly affair. In this paper we have presented a framework for semi-automatic construction of ontology using knowledge discovery techniques with an aim to reduce the effort required to produce ontology for a new domain.

The main contribution of this research paper is the concept of using both Web Content Mining and Web Usage Mining for semi-automatically developing the ontologies. The main idea behind this concept was to incorporate the Web author's ideas as well as Web users' intentions in the ontology development process. However, reliability of using only Web Usage Mining information was seen as not a viable solution because of the rapid changing nature seen in some of the web sites and hence, identifying users' navigational patterns is difficult and it may not reveal adequate information. Therefore, the proposed methodology extracts concepts and conceptual relationships using Web Content Mining and the information revealed by Web Usage Mining is incorporated to refine the ontologies. The quality of the constructed Ontology is assessed using the User Evaluation method. In future research, we plan to combine Web Structure Mining with the proposed approach in developing ontologies, which could be expected to give the profitable results. Another direction in which, the proposed approach can be extended is, by applying Web Mining techniques to domain specific multiple Web sites to develop domain ontology.

REFERENCES

[1] Maedche and S. Staab, (2001). "Ontology Learning for the Semantic Web ", IEEE Intelligent Systems, *Special Issue on the Semantic Web*, Vol.16, No.2, pp. 72 - 79.

[2] G.Stumme , A. Hotho and B. Berendt (2006). "Semantic Web Mining: State of the art and future directions", *Journal of Web Semantics: Science, Services and Agents on the World Wide Web*, Vol. 4, No. 2, pp. 124-143.

[3] Bettina Berendt, Andreas Hotho, and Gerd Stumme (2002). "Towards semantic web mining ", In *International Semantic Web Conference (ISWC, Springer)*, pp. 264-278.

[4] G.Stumme, B. Berendt and A Hotho, (2004). "Usage Mining for and on the Semantic Web ", *Data Mining: Next Generation Challenges and Future Directions*", *AAAI*/MIT Press, pp. 461-480.

[5] Julia Hoxha, Martin Junghans, Sudhir Agarwal. "Enabling semantic analysis of user browsing patterns in the Web of Data ", In *Proceedings of the 2nd International Workshop on Usage Analysis and the Web of Data (USEWOD), 21st International World Wide Web Conference* (WWW 2012), Arxiv abs/1204.2713, Lyon, France, April, 2012.

[6] T. Gruber (1994)."Towards principles for the design of ontologies used for knowledge sharing ", *International Journal of Human and Computer Studies*, Vol. 43, pp. 907–928.

[7] Noy, N. F., Sintek, M., Decker, S., Crubezy, M., Fergerson,R.W., and Musen, M.A. (2001). "Creating Semantic Web Contents with Protege-2000 ", In *IEEE Intelligent Systems*, Vol. 16, No. 2, pp. 60-71.

[8] Sure, Y., Erdmann, M., Angele, J., Staab, S., Studer, R., and Wenke, D. (2002). "OntoEdit: Collaborative ontology development for the Semantic Web ", In *International Semantic Web Conference (ISWC 2002)*, Sardinia, Italy.

[9] Gomez-Perez, A., Manzano-Macho, D. (2003). "OntoWeb Deliverable 1.5: A Survey of Ontology Learning Methods and Techniques ", *Universidad Politecnica* de Madrid.

[10] Shamsfard, M. and Barforoush, A. A. (2003). "The state of the art in ontology learning: A framework for comparison ", *The Knowledge Engineering Review*, Vol. 18, No.4, pp. 293-316.

[11] Sabou, M., Wroe, C., Goble, C., and Mishne, G.(2005). "Learning Domain Ontologies for Web ServiceDescriptions: an Experiment in Bioinformatics", In *Proceedings of the 14th International World Wide Web Conference (WWW2005)*, Chiba, Japan.

[12] Maedche and S. Staab (2000). "Semi-automatic engineering of ontologies from text", In *Proceedings of the 12th Internal Conference on Software and Knowledge Engineering*, pp. 231-239.

[13] Sean P. Igo, E. Desmontils, and Ellen Rilo, (2009). "Corpus-based semantic lexicon induction with web-based collaboration ", In *NAACL-09 Workshop on Unsupervised and Minimally Supervised Learning of Lexical Semantics*.

[14] Eduard Hovy, Zornitsa Kozareva, and Ellen Rilo, (2009). "Learning and evaluating the content and structure of a term taxonomy ", In *Proceedings of AAAI-09 Spring Symposium*.

[15] D. P. T. Nguyen, Y. Matsuo, and M. Ishizuka,(2007). "Exploiting Syntactic and Semantic Information for Relation Extraction from Wikipedia ", *IJCAI Workshop on Text-Mining Link-Analysis*.

[16] Changqing Li and Tok Wang Ling, (2005). "From XML to Semantic Web", In *10th International Conference on Database Systems for Advanced Applications*, pp. 582-587.

[17] Mikroyannidis and B. Theodoulidis, (2005). "Web Usage Driven Adaptation of the Semantic Web ", In *Proc. End User Aspects of the Semantic Web Workshop, 2nd European Semantic Web Conference (ESWC 2005)*, Heraklion, Greece, pp. 137-147.

[18] Jayatilaka A.D.S and Wimalarathne G.D.S.P, (2011). "Knowledge Extraction for Semantic Web Using Web Mining ", *International conference on Advances in ICT for emerging regions (ICTER 2011),* pp. 89-94.

[19] Tatyana Ivanova, (2010). "A Semi-Automatic Ontology Learning Method for E-Learning Resources Terminology Extraction", *International Conference on ICL*, pp. 1030-1034.

[20] Marko Brunzel and Myra Spiliopoulou, (2006). "Discovering multi terms and co-hyponymy from XHTML documents with xtreem ", In *Proceedings of PAKDD Workshop on Knowledge Discovery from XML Documents (KDXD 2006)*, LNCS 3915.

[21] Hazman, M., El-Beltagy, S. R., and Rafea, A. (2009). "Ontology Learning from Domain Specific Web Documents ", *In International Journal of Metadata, Semantics and Ontologies,* Vol. 4, No. 1-2, pp. 24 – 33.

[22] R. Agrawal and R. Srikant. (1995). "Mining Sequential Patterns ", *Proceedings. of the 11th Int'l Conference on Data Engineering (ICDE- 95)*, pp. 3-14.

[23] Yan, X., Han, J., Afshar, R. (2003). "CloSpan: Mining Closed Sequential patterns in large datasets ", In *Proceedings of SIAM International Conference on Data Mining*, pp.166–177.

[24] J. Srivastava, R. Cooley, M. Deshpande and P. Tan, (2000). "Web usage mining: Discovery and applications of usage patterns from Web data ", *SIGKDD Explorations*, Vol. 1, No. 2, pp. 12-23.

[25] http://protege.stanford.edu

[26] Brank, J., Grobelnik, M., and Mladenic (2005). A survey of ontology evaluation techniques ", In *Proceedings of the 8th International Multi-Conference Information Society IS-2005*.

AUTHORS

Mr. C. Ramesh is working as Associate professor in Computer Science and Engineering Department at CVR College of Engineering. He is pursuing his PhD from Jawaharlal Nehru Technological University, Hyderabad. He received his B.E in Computer Science and Engineering from Osmania University and M.Tech from Jawaharlal Nehru Technological University, Hyderabad. He has guided 16 M.Tech projects and published 3 papers in International and National journals.

His areas of interest include Databases, Data Mining, Semantic Web Mining, Web Usage Mining and Social Networks. He has 14 years of teaching experience.

K.V. Chalapati Rao is a Professor of Computer Science & Engineering and Dean, Academics at CVR College of Engineering. Prior to joining the CVR, he served Osmania University as a Professor and Head, Department of CSE and Dean of Engineering. After obtaining his PhD, he joined Electronics Corporation of India Limited and worked in various capacities for 16 years, before joining the Osmania University.

He guided number of PhD scholars in areas of Real time systems, Operating Systems, Software Engineer. He has published a number of papers in International/National journals/conferences including IEEE, ACM, Springer and Elsevier.

A.Govardhan is presently a Professor of Computer Science and Engineering, Director at School of Information Technology and Executive Council Member, Jawaharlal Nehru Technological University Hyderabad (JNTUH), India. He served and held several Academic and Administrative positions including Director of Evaluation, Principal, Head of the Department, Chairman and Member of Boards of Studies and Students' Advisor. He received Ph.D. degree in Computer Science and Engineering from Jawaharlal Nehru Technological University in 2003, M.Tech from Jawaharlal Nehru University in 1994 and B.E from Osmania University in 1992.

He is the recipient of 24 International and National Awards. He has guided 48 Ph.D theses, 1 M.Phil, 135 M.Tech projects and he has published 350 research papers at International/National Journals/Conferences including IEEE, ACM, Springer and Elsevier. He is a Member on Editorial Board for 11 International Journals and PC Member and Reviewer for several International/National Conferences. He is a Life Member/Member in several Professional and Service Oriented Bodies. His research of interest includes Databases, Data Warehousing & Mining, Information Retrieval, Computer Networks, Image Processing, Software Engineering and Object Oriented Technologies.

INTENTIONAL BLANK

COOPERATIVE DATA SHARING WITH SECURITY IN VEHICULAR AD-HOC NETWORKS

Deepa B[1] and Dr. S A Kulkarni[2]

[1]IV Sem M. Tech, Dept of CSE, KLS Gogte Institute of Technology, Belagavi
deepa.bangarshetru@gmail.com
[2]Professor and HOD of ISE, KLS Gogte Institute of Technology, Belagavi
shri1_kulkarni@yahoo.com

ABSTRACT

Vehicles download the data when passing through a drive through the road (RSU) and then share the data after travelling outside the coverage of RSU.A key issue of downloading cooperative data is how effectively data is shared among them self. Developing an application layer data exchange protocol for the coordination of vehicles to exchange data according to their geographic locations. Coordinated sharing can avoid medium access control (MAC) layer collisions and the hidden terminal effect can be avoided in the multi-hop transmission. A salient feature of the application layer data exchange protocol, in the voluntary services, Vehicles purchase the requested data from service provider via RSUs. In this project, we propose a cooperative data sharing with secure framework for voluntary services in special vehicles networks (VANETs). We also concentrate on security in the process of downloading data and sharing. Applicants to ensure exclusive access to data applied and security of the vehicles involved in the implementation.

KEYWORDS

VANET, RSU, Security, Data Sharing & Voluntary Services.

1. INTRODUCTION

VANET- Vehicular networks is likely to develop in the upcoming years and thus become the most applicable form of ad hoc networks. Vehicular Ad hoc Network (VANET) consists of the imperative elements of Intelligent Transportation System (ITS) in which vehicles are arranged with several short-range and medium-range wireless communications. In VANET two kind's communication are possible. One is vehicle-to-vehicle (V-2-V) communication; the other is roadside-to-vehicle communications (V-2-R).By V-2-V communication, people can obtain more information and use the shared information to improve road safety. By V-2-R communication, people can communicate with RSU to access internet for downloading and updating files or inquire neighbourhood location information. Thus, compared with the traditional pure infrastructure-based network, the hybrid of V-2-V and V-2-R communications is promising since

David C. Wyld et al. (Eds) : ACITY, DPPR, VLSI, WiMNET, AIAA, CNDC - 2015
pp. 53–62, 2015. © CS & IT-CSCP 2015 DOI : 10.5121/csit.2015.51305

it can not only overcome the disadvantages of infrastructure-based network, but can also overcome the disadvantage of non-infrastructure-based network.

In recent years, VANETs- vehicular ad hoc networks have gained much attention in the world of automobiles and Research. One reason is interest in an increasing number of applications designed for safety of passengers such as traffic jam detection and cooperative driving and also for emergency braking, As well as in applications for the comfort of passengers like games, chat-rooms and distribution of vehicle data (eg CarTorrent). The increased use of software has not only affected the automotive guarantee costs, but has also made most difficult to car repairs. Data downloading is a practical and prominent application in VANETs-vehicular ad hoc networks, which can bring comfort and entertainment to users. In data downloading, vehicles send service requests and then get the data stream from the current or the next roadside units (RSU). In the downloading application, the amount of data a vehicle can download at a drive-through of a RSU is very limited due to the short connection time. Cooperative download is a promising scheme in which vehicles download the data when passed through a RSU and then share data when traveling outside the scope of the communications of RSU. Thus, the total amount of data that can download a particular vehicle will increase.

A key issue in cooperative download is how vehicles share data with others. There are some existing studies on data exchange in VANETs [1]. However, existing exchange protocols are limited to issues of medium access control (MAC) layer of the collisions, limited applicability to multiple data exchange units, and there is no guarantee of receipt of complete data.

We propose an application layer protocol for data exchange with the assumption that each vehicle knows the positions of the own and neighboring vehicles (which can be obtained through global positioning system (GPS) and related security messages transmitted regularly by neighboring vehicles [2]). In the proposed protocol, vehicles used for coordination channel to coordinate relay transmissions in VANETs for data exchange based on GPS vehicle location. With such cooperative exchange, collisions and MAC layer hidden terminal effect can be avoided in the data channel. In addition, Design a stylish selection of relay vehicles mechanism for the space between the two RSU can be completely exploited for data exchange. A prominent feature of the proposal exchange protocol is that it can ensure the receipt of the data for each applicant vehicle pass an RSU. Security is also critical issue.

Characteristics of voluntary services require exclusive access to applicant's data.

In summary, develop a framework for cooperative secure data sharing with the following contributions.

1. Designed an application layer protocol for data exchange to facilitate data sharing with the coordinated transmission. With such coordinated sharing can avoid medium access control (MAC) layer collisions and the hidden terminal effect can be avoided in the multi-hop transmission.
2. Security protocol for voluntary services VANETs are developed, which can ensure applicants exclusive access to data applied and also ensures security of the vehicles involved in the implementation.

1.1 Features and challenge

To develop this system we come across some of the difficulties

- Vehicles are moving faster and therefore lifetimes of the communication links are shorter; therefore, the links are facing rapid changes in the network topology.
- Vehicles are high mobile and are usually constrained with layout of the road, speed limits, traffic and vehicle destination. If the vehicle is in exceeding the speed limit then it results in receiving an incomplete data. This requires the intelligent file transfer.
- The existing sharing protocols constrained with issues of medium access control (MAC) layer collisions, limited applicability to sharing multiple data units, and no guarantee of complete data receiving.

2. LITERATURE SURVEY

S. Ahmed and S. S. Kanhere in [1] proposed a co-operative content distribution scheme based on novel network coding called VANETCODE for Content distribution in Vehicular Ad-Hoc Networks (VANET) is challenging due to the high mobility, rapidly topology changing and intermittent connectivity observed in these type of networks. Using VANETCODE, leverages of the wireless medium of the broadcasting nature to accelerate the distribution of encoded blocks amongst neighboring one-hop neighbors and is completely independent of routing.

X. Lin et. al proposed a secure data downloading protocol with preserving privacy in VANETs - vehicular ad hoc networks [2]. Which Enables vehicles to download data from RSUs Securely With Their privacy protection under one or Even When multiple RSUs are compromised .This protocol give the guarantees for vehicles to exclusive access their requested data while eavesdroppers cannot obtain any private information of the vehicles.

K. Sampigethava et. al describe techniques used for privacy-preserving and secure protocol based on identity (ID)-based and group signature [3]. This protocol gives the guarantee the requirements of privacy and security of the each vehicle. But it can also give the each vehicle desired traceability in the event that the ID of the sender message must be revealed by the authority for any disputes of an event.

Y. Hao et. al give details of how the problem involved in controlling of unauthorized vehicle tracking based on their broadcast communications media, in order to improve user location privacy in VANET. [4] AMOEBA provides location privacy by utilizing the location group navigation of vehicles. By using vehicle groups for anonymous access to applications location-based services in VANET, for privacy protection of user. The robustness of privacy of the user provided is considered under various attacks.

J. Byers et. al describe fully scalable and ideal protocol for the applications such as reliable distribution of bulk data for that we call a digital fountain [5]. In this many number of heterogeneous receivers at times of their choice to procure content with maximum efficiency. Here, no feedback channels are required in order to ensure the reliable delivery, even when the face of high loss rates.

3. SYSTEM ARCHITECTURE

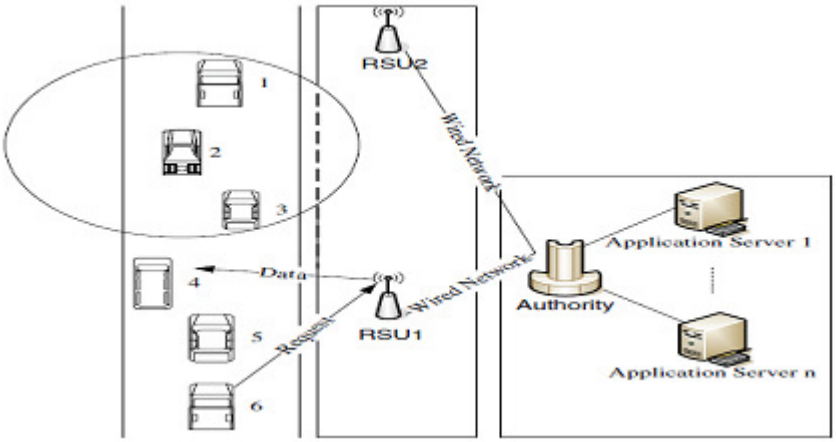


Figure 1: System Architecture

Applicant vehicle download the data from the application server via RSU. Initially applicant vehicle send request for data to RSU, then RSU will forward that request to authority. Authority is responsible for selection of downloading vehicle based on geographic information and also check the identity of the vehicle whether it is valid to purchase the data or not from application server. If vehicle is valid to purchase the data then it will download the data from the server and that will send to RSU. Finally RSU sends data to the Applicant vehicle.

The application authority and application servers: These are responsible for the management and provision of service data, respectively. The authority knows all the keys and is in charge of programming service. They can be kept either by the authority or third party operators.

Road side infrastructure: Consisting of RSUs deployed at the edges of the roads that are responsible for forwarding request and response. RSUs communicate with authority via wired network.

Nodes: These are ordinary vehicles on the street and highway road that can communicate with each other and RSUs through radio.

3.1 Vehicle Classification

Classifies the vehicle into three categories i.e Applicant Vehicle and Downloading Vehicle and Relay Vehicle.

Applicant Vehicle: These are the vehicle they are likely to purchase the data.
Downloading Vehicle: These vehicle download the data from the RSU for applicants. These are assigned by the authority according to geographic positions.
Relay Vehicle: These are responsible for forwarding data to buyers which are more than one hop away from downloading vehicles

4. ANALYSIS AND RESULTS

The experiment was conducted for 50, 55, 60, so on number of sent packets. The comparison between ADOV & ADOV_DES is made. In figure 6 we show the transmission of message to the v which is conducted in NS-2 and also shows the transmitted packets to vehicles from RSU (Road Side Unit). The RSU broadcasts the message any vehicle in the range will receive the message. The message transmission to the nearest vehicle, Without providing security and with providing security is shown in the following figures.

Title:-Run simrun.tcl for 50 packets, Without providing security (AODV) total number of sent packets will be 50 and number of received packets will be 49 and with delay was 13.0747. Consider 73 as applicant vehicle receives data from application server via RSU (67) are shown in the figure 2 and figure 3.

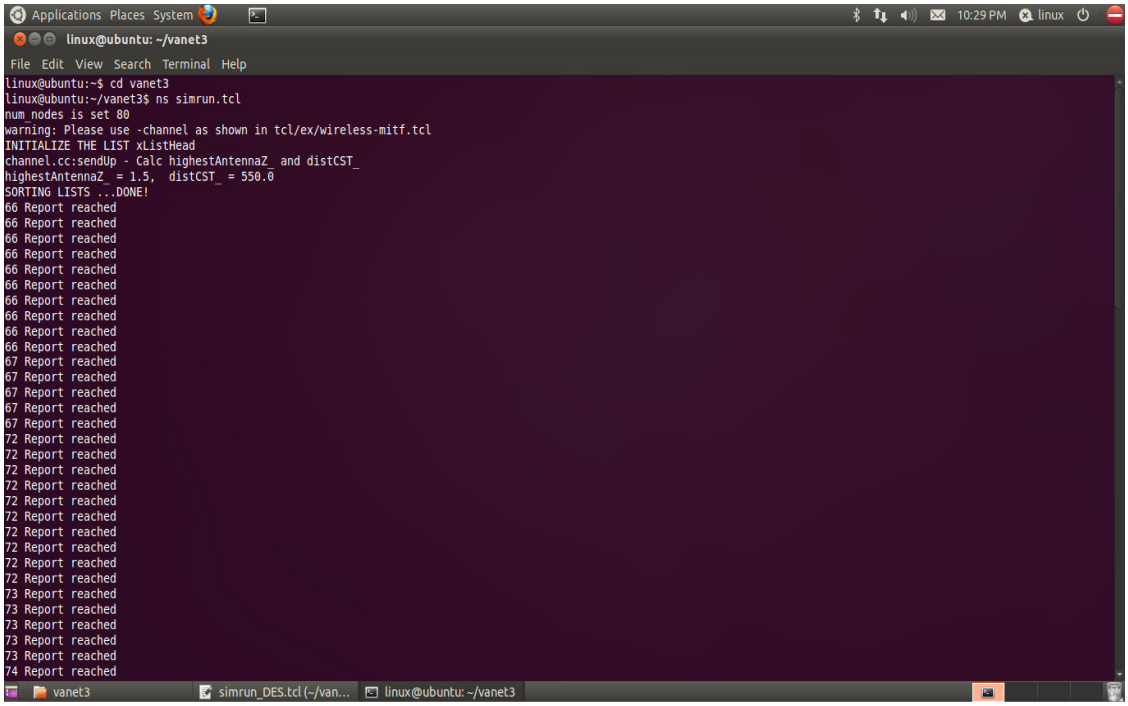


Figure 2: Run simrun.tcl for 50 packets

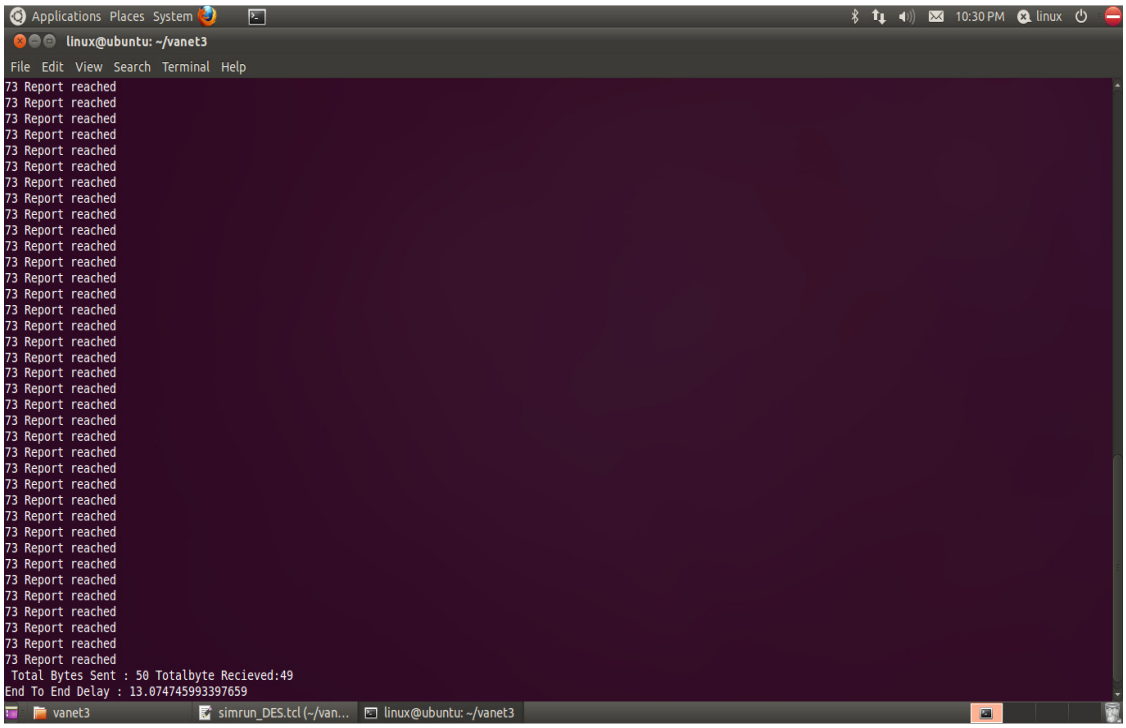


Figure 3: Run simrun.tcl for 50 packets

Title:-NAM Window with all the nodes before starting. Create road mapping using 65 nodes. Randomly use 72-75 nodes are applicants. Applicants are the vehicles to purchase the data from application server via RSU's.Application server1 located at position (400,610) sends data to Authority located at position (450,630).

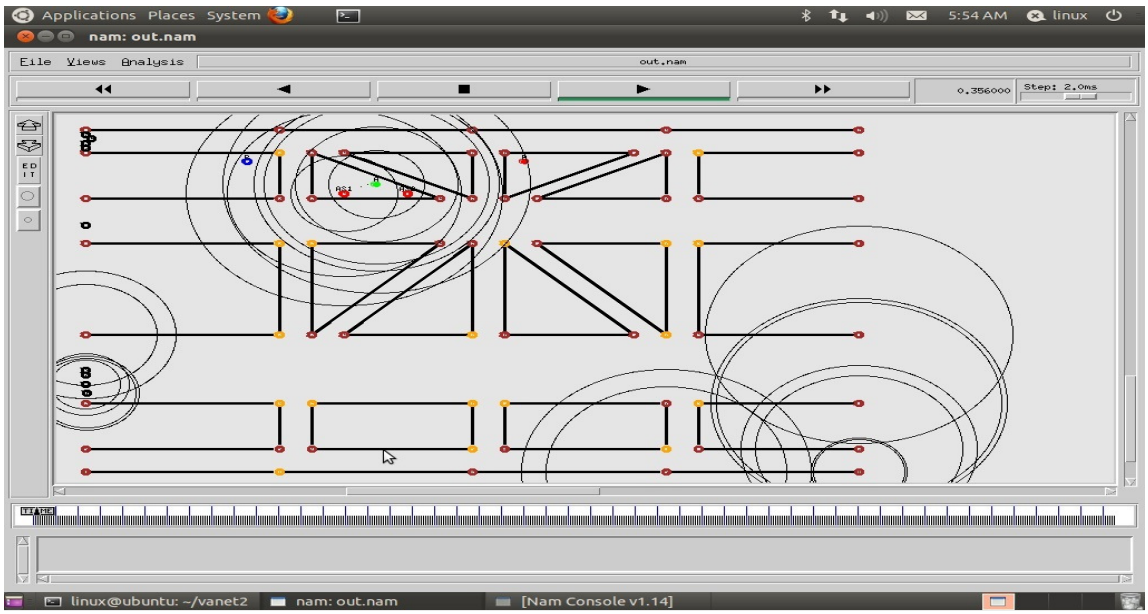


Figure 4: Application Server1 sends data to authority

Title:-Authority located at (450,630) sends data to RSU (67) located at (250,680).

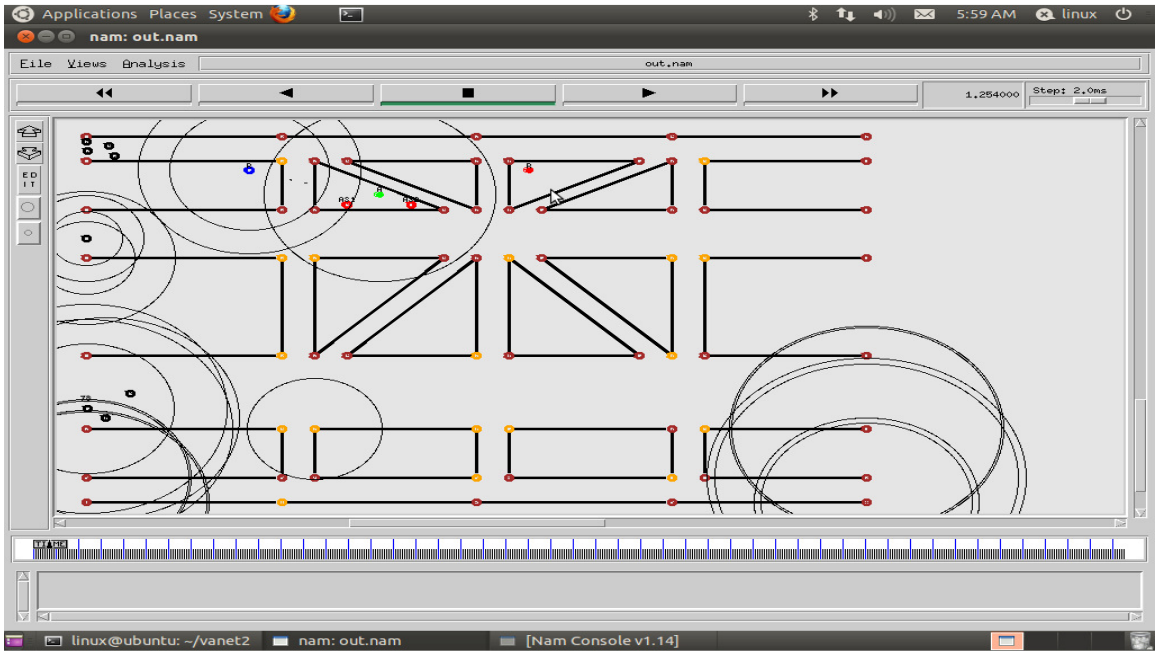


Figure 5: Authority sends data to RSU

Title:- RSU (67) located at (250,680) sends data to applicant vehicle (73) without providing security.

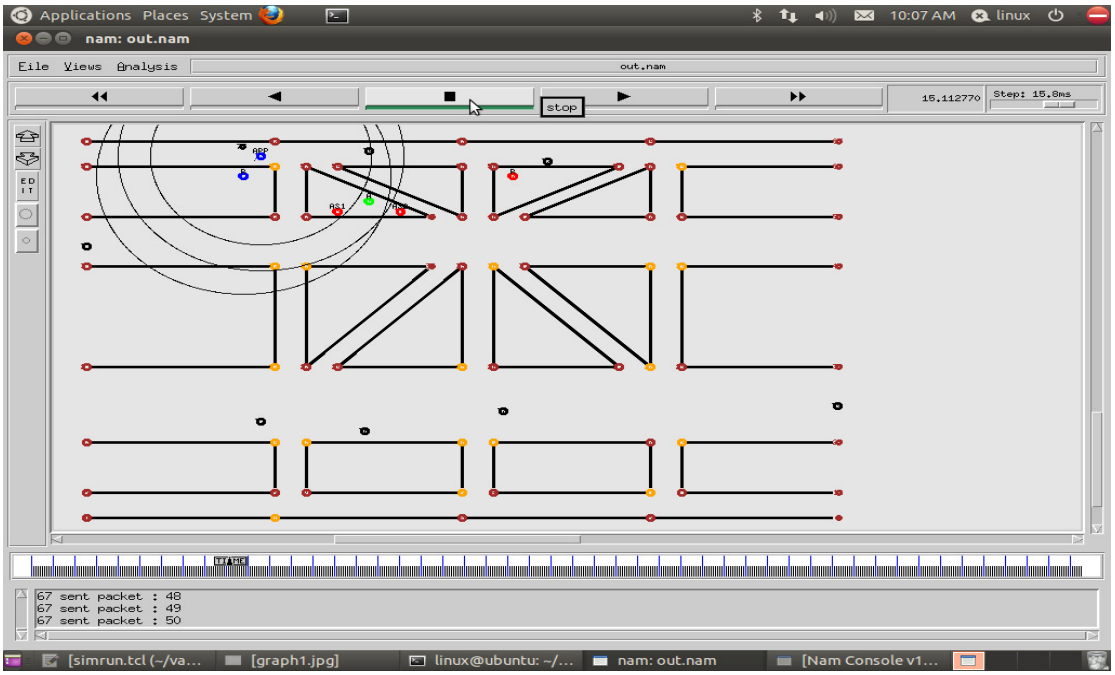


Figure 6: RSU sends data to applicant vehicle

Title:-Run simrun_DES.tcl for 50 packets, With providing security (AODV_DES) total number of sent packets will be 50 and number of received packets will be 49 and with delay was 14.047. Consider 75 as Applicant vehicle receives data from application server via RSU's after receiving data, the data will be decrypted and also cross verifies the signature, if signature is matched than message is accepted else message will be rejected are shown in the figure 7 to figure 9.

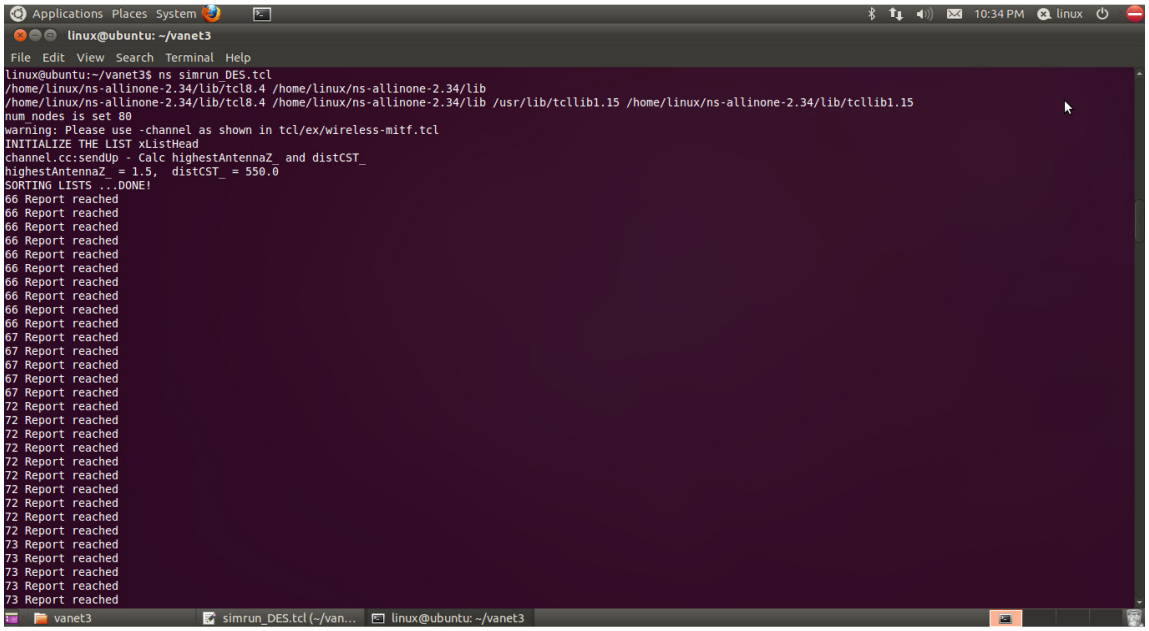


Figure 7: Run simrun_DES.tcl for 50 packets

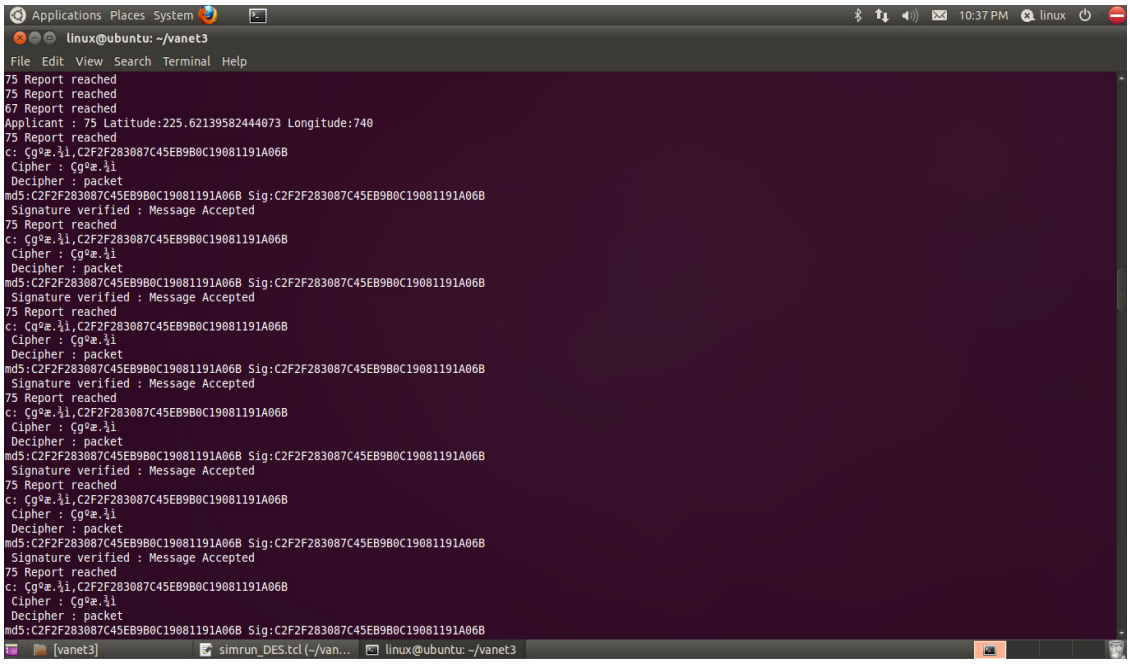


Figure 8: Run simrun_DES.tcl for 50 packets

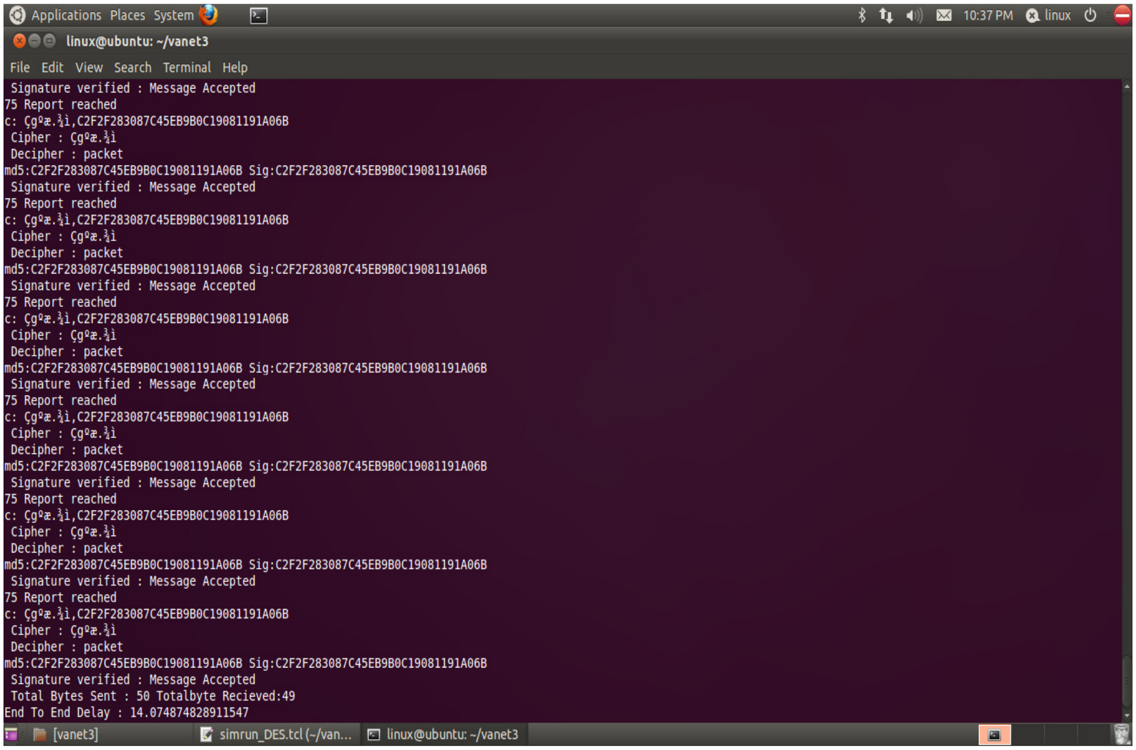


Figure 9: Run simrun_DES.tcl for 50 packets

Title:- RSU (67) located at (250,680) sends the packets to applicant vehicle (75) with providing security.

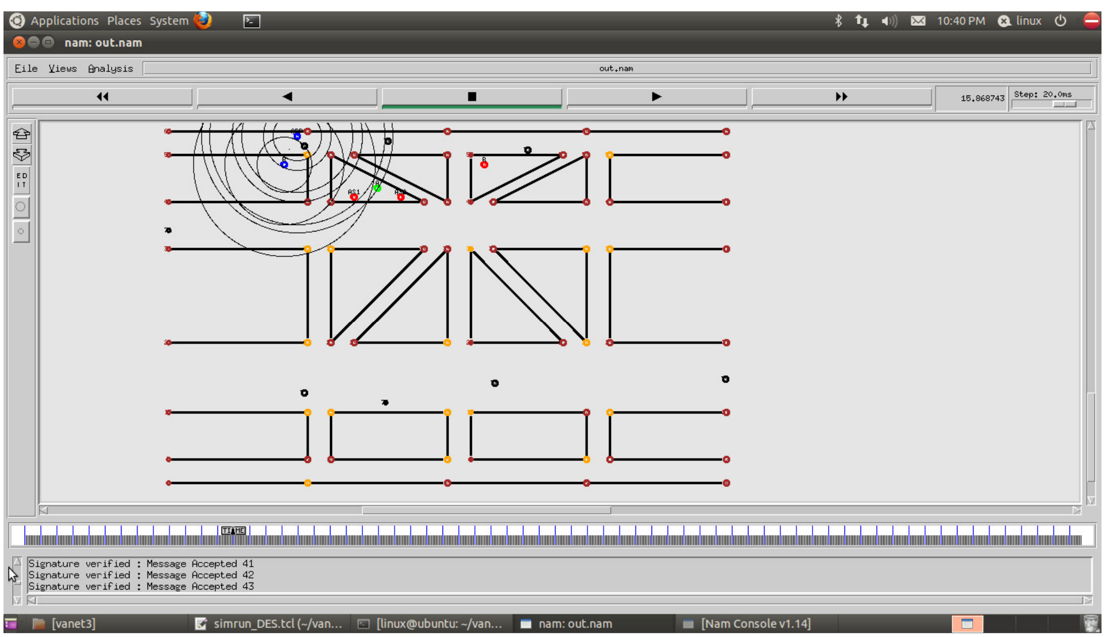


Figure 10: RSU sends packets to applicant vehicle

In Figure 11 consider three samples, sample1: Describes total number of sent packets is 50 and received 49 and dropped is 1. And sample2: Describes total number of sent packets is 55 and received 54 and dropped is 1.And sample3: Describes total number of sent packets is 60 and received 59 and dropped is 1.By observing the below chart we conclude that; the complete data with 1 dropped packet is received.

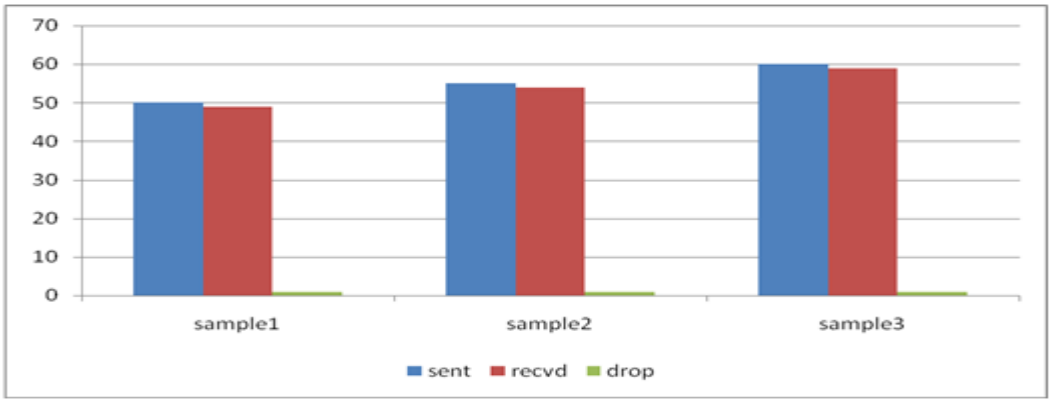


Figure 11: The comparison of delay of packets

5. CONCLUSION

Based on the above experiments & results conducted, we infer that; the transmission with DES & MD5 algorithm is secured & efficient. With such coordinated sharing can avoid medium access control (MAC) layer collisions and the hidden terminal effect can be avoided in the multi-hop transmission and also provides security protocol for voluntary services VANETs, which can ensure security and improve the efficiency of developed framework.

In future we want to share the messages using different techniques while with security as main concern & in real time transmission.

REFERENCES

[1] S. Ahmed and S. S. Kanhere, "VANETCODE: Network coding to enhance cooperative downloading in vehicular ad hoc networks," in Proc. IWCMC, 2006.

[2] X. Lin, X. Sun, P.-H. Ho, and X. Shen, "GSIS: A secure and privacy preserving protocol for vehicular communications," IEEE Trans. Veh. Technol., vol. 56, no. 6, pp. 3442–3456, 2007.

[3] K. Sampigethava, M. Li, L. Huang, and R. Poovendran, "AMOEBA:Robust location privacy scheme for VANET," IEEE J. Sel. Areas Commun., vol. 25, no. 8, pp. 1569–1589, 2007.

[4] Y. Hao, J. Tang, Y. Cheng, and C. Zhou, "Secure data downloading with privacy preservation in vehicular ad hoc networks," in Proc. IEEE ICC, May 2010.

[5] J. Byers, M. Luby, M. Mitzenmacher, and A. Rege, "A digital fountain approach to reliable distribution of bulk data," in Proc. ACM SIGCOMM,1998, pp. 56–678.

[6] F. Ye, S. Roy, and H. Wang, "Efficient data dissemination in vehicular ad hoc networks," IEEE J. Sel. Areas Commun., vol. 30, no. 4, pp. 769–779, May 2012.

[7] S. Lee, G. Pan, J. Park, M. Gerla, and S. Lu. "Secure incentives for commercial ad dissemination in vehicular networks," in Proc. ACM MobiHoc, 2007.

STABLE DRUG DESIGNING BY MINIMIZING DRUG PROTEIN INTERACTION ENERGY USING PSO

Anupam Ghosh[1], Mainak Talukdar[2] and Uttam Kumar Roy[3]

[1]Indian Institute of Technology Bombay, Mumbai, India
anupam.ghsh@gmail.com
[2]Cognizant Technology Solutions, Kolkata, India
codes.mnk@gmail.com
[3]Jadavpur University, Kolkata, India
royuttam@gmail.com

ABSTRACT

Each and every biological function in living organism happens as a result of protein-protein interactions. The diseases are no exception to this. Identifying one or more proteins for a particular disease and then designing a suitable chemical compound (known as drug) to destroy these proteins has been an interesting topic of research in bio-informatics. In previous methods, drugs were designed using only seven chemical components and were represented as a fixed-length tree. But in reality, a drug contains many chemical groups collectively known as pharmacophore. Moreover, the chemical length of the drug cannot be determined before designing the drug.

In the present work, a Particle Swarm Optimization (PSO) based methodology has been proposed to find out a suitable drug for a particular disease so that the drug-protein interaction becomes stable. In the proposed algorithm, the drug is represented as a variable length tree and essential functional groups are arranged in different positions of that drug. Finally, the structure of the drug is obtained and its docking energy is minimized simultaneously. Also, the orientation of chemical groups in the drug is tested so that it can bind to a particular active site of a target protein and the drug fits well inside the active site of target protein. Here, several inter-molecular forces have been considered for accuracy of the docking energy. Results show that PSO performs better than the earlier methods.

KEYWORDS

Active Site, Docking, Electrostatic Force, Proteins, Van Der Waals Force

1. INTRODUCTION

Protein is a macro molecule primarily consisting of amino acids [1]. Protein is highly responsible for structural and functional characteristics of cells, and communication of biological signals among cells. All proteins available in the nature are not useful for the living organism. However, every biological function in living organism happens as a result of protein-protein interactions [2,

David C. Wyld et al. (Eds) : ACITY, DPPR, VLSI, WiMNET, AIAA, CNDC - 2015
pp. 63–74, 2015. © CS & IT-CSCP 2015 DOI : 10.5121/csit.2015.51306

3]. There are evidences of proteins which cause fatal or infective diseases. Researchers take keen interest to identify appropriate drug that fits well in the active sites of harmful protein. These drugs, which can bind in the active site of target protein and therefore change the functional behavior of that particular protein, are called *ligands*. This mechanism of binding of drug with the target protein is called *docking* [4]. The challenge is to predict an accurate structure of the ligand when the active site configuration of the target protein is known.

Docking is very much important in cellular biology. In docking, proteins interact with themselves and with other molecular components. It is the key concept to oriental drug design. The result of docking can be used to find inhibitors for specific target proteins and thus to design new stable drugs. Protein-ligand docking is an energy minimization search and optimization problem with the aim to find the best ligand conformation for active sites of a target protein.

In this paper, the scope of the well-known Particle Swarm Optimization (PSO) algorithm [5, 2, 6, 7] is studied to determine the ligand structure to be docked at the active site of the target protein. The choice of PSO in the present context is inspired by social behavior of bird flocking and fish schooling [8]. It begins with a random population and then searches for optimal solution by updating the population based on fitness of the current solution.

Evolutionary computation is used to place functional groups in the nodes of the variable tree structured ligand. The tree structure helps in connecting primitive fragments or radicals to determine the right candidate solution for the ligand that best suits with the active site of the protein. The PSO based algorithm flows through the entire search space, and records the best individuals they have met. At each step, it changes its position according to the best individuals to reach a new position. In this way, the whole population evolves towards the optimum, step by step.

A ‘fitness function’ is generally introduced in a meta-heuristics algorithm to determine the quality of the desired solutions for an optimization problem. Naturally, the better the formulation of the fitness function, the better is the expected quality of the trial solutions. In the present context of the ligand-docking problem, optimal selection of the ligand is inspired by minimization of an energy function that determines the stable connectivity between the protein and the ligand. So, the fitness function here is an energy function, whose minimization yields trial solutions to the problem. One important factor in the field of drug design is the identification of proper ligand. In general, one or more proteins are typically involved in the bio-chemical pathway of a disease. The treatment aims to appropriately identify those proteins and reduce their effects by designing a ligand molecule [9] that can bind to protein’s active site. That is, the structure of a ligand molecule is evolved from a set of groups in close proximity to crucial residues of the protein; a molecule is thereby designed that fits the protein target receptor such that a criterion for Van Der Waals & electrostatic interaction energy is optimized. We propose to adopt this approach in this paper.

In this paper, we propose a novel approach by which the drug is represented by a variable length tree-like structure, forty-five functional groups, Van Der Waals as well as electrostatics force to design the ligand structure. We have applied our approach to Human Rhinovirus stain 14, Plasmodium Falciparum and HIV-1 protease viruses. We have significantly improved the work done by earlier researchers due to following reasons 1) A new representation for the ligand is used; 2) Both Van Der Waal and electrostatic energies are optimized and 3) PSO is used because of its ability to handle to the ligand design problem. It can be seen that PSO method has performed quite satisfactorily in comparison with Genetic Algorithm (GA) (fixed length tree)

[10], Neighbourhood Based Genetic Algorithm (NBGA) (fixed length tree) [11], and variable length Neighbourhood Based Genetic Algorithm (VNBGA) (variable length tree) [12].

The rest of the paper is organized as follows. Section 2 briefs the related work. Section 3 contains the formulation of protein–ligand docking problem; section 4 depicts the principles used to predict the ligand structures. In section 5, PSO algorithm used to find the best ligand structure is described. The pseudo-code for solving the given constrained optimization function is given in section 6. A comparison of experiment results for 3 proteins in section 7 is also given. Section 8 concludes the paper with the future work.

2. RELATED WORK

A fixed length Genetic Algorithm approach for ligand design was used in [10, 8, 13, 14] to evolve molecular structure of possible ligand that binds to a given target protein. The drug is represented by a fixed tree-like structure, comprising of molecules at the nodes and the bonds as links. Evidently, an a priori knowledge of the size and length of the tree is difficult to obtain before the experiment.

Another approach for ligand design, which is based on variable length representation of trees on both sides of the pharmacophore was studied by Bandopadhyay et al [12]. However, the approach is restricted to build the ligand in two-dimensional space from a small suite of seven functional groups. Furthermore, the fitness value of the ligand is confined to Van Der Waals force only.

One more approach for ligand design, that is based on the presence of a fixed pharmacophore and that uses the search capabilities of genetic algorithms, was studied by Goah and Foster [10], where the harmful protein human Rhinovirus strain14 was used as the target. This pioneering work assumed a fixed tree structure representation of the molecule on both sides of the pharmacophore. Evidently, an a priori knowledge of the size of the tree is difficult to obtain. Moreover, it is known that no unique ligand structure is best for a given active site geometry.

Furthermore, few more algorithms were designed by researchers for ligand design *viz.* AutoDock 3.05, SODOCK and PSO@AUTODOCK. In [6], a novel optimization algorithm SODOCK is used for solving protein–ligand docking problems. And results are compared with the performance of AutoDock 3.05 [15]. SODOCK is based on particle swarm optimization and a local search strategy. The work uses the environment and energy function of AutoDock 3.05. Results show that SODOCK performs better than AutoDock in terms of convergence performance, robustness, and obtained energy.

In [16], a novel meta-heuristic algorithm called PSO@AUTODOCK based on varCPSO and varCPSO-ls is used to calculate the docking energy. varCPSO stands for velocity adaptive and regenerative CPSO. varCPSO-ls is varCPSO with local search technique to avoid convergence at local minima. In addition to that, comparison is done with other docking techniques as GOLD 3.0, DOCK 6.0, FLEXX 2.2.0, AUTODOCK 3.05, and SODOCK. It has been shown that PSO@AUTODOCK gives highly efficient docking program in terms of speed and quality for flexible peptide–protein docking and virtual screening studies.

In our paper, we propose a PSO based approach wherein a variable length tree-like structure, forty-five functional groups, Van Der Waals and electrostatics force is used to design the ligand structure. We have compared our PSO based approach with GA (fixed length tree), NBGA (fixed length tree) and VNBGA (variable length tree).

3. FORMULATION OF THE PROBLEM

In protein-ligand docking problem, the objective is to minimize the energy. Firstly, the internal energy of the ligand should be minimized for better stability of the ligand [17]. The inter-molecular energy value, which is thereafter optimized, is the interaction energy between the ligand and the active site of the receptor protein. This energy calculation is based on the proximity of the different residues in the active site of the receptor protein to the closest functional groups in the ligand and their chemical properties. The inter-molecular interaction energy is computed in terms of the Van Der Waals energy and the electrostatic energy.

It is noted that the distance between a residue of the target protein receptor and its closest functional group should be between $0.7\overset{0}{A}$ and $2.7\overset{0}{A}$. If the functional group and closest residue are of different polarity, then electrostatic force of attraction also acts between those molecules. The interaction energy of the ligand with the protein is the sum of Van Der Waals Force and electrostatic energy [4].

$$\boldsymbol{E}_{electrostatics} = q_A q_B \Big/ 4\pi\varepsilon_0 r_{AB} \qquad (1)$$

Value of $1\Big/4\pi\varepsilon_0 = 9*10^9 N - m^2 / c^2$

Where q_A and q_B are the charges of the two atoms is the separation, ε is the dielectric constant of the surrounding medium $\boldsymbol{r}$ is the distance between charges.

$$E_{vdw} = A\Big/r^{12} - B\Big/r^{6} \qquad (2)$$

Where A and B are constants and $\boldsymbol{r}$ is distance between the molecules. Value of A and B depends on the atom pair.

$$\boldsymbol{E}_{total} = \boldsymbol{E}_{electrostatics} + \boldsymbol{E}_{vdw} \qquad (3)$$

The fitness value is taken as $F = 1\Big/E_{total}$

4. FORMATION OF A LIGAND

In the proposed work, we consider that the ligands are built using the fragments from the suite as mentioned in Figure 4. Proteins with known active site configuration are used for evolving ligand structures. The specific target is the known antiviral binding site of the Human Rhinovirus strain 14, Plasmodium Falciparum and HIV-1 Protease. The active site of Human Rhinovirus strain 14 is known as the VP1 barrel as it resembles with a barrel. The molecules which can easily be fit in the structure having minimum interaction energy will be the evolved drug (i.e. ligand). For simplification; a 2-dimensional structure is chosen.

Figure 1 illustrates the active site of Human Rhinovirus strain 14, Figure 2 illustrates the active site of Plasmodium Falciparum and Figure 3 illustrates the active site of HIV-1 protease. For designing the ligand, the co-ordinates of the active site of the protein must be known.

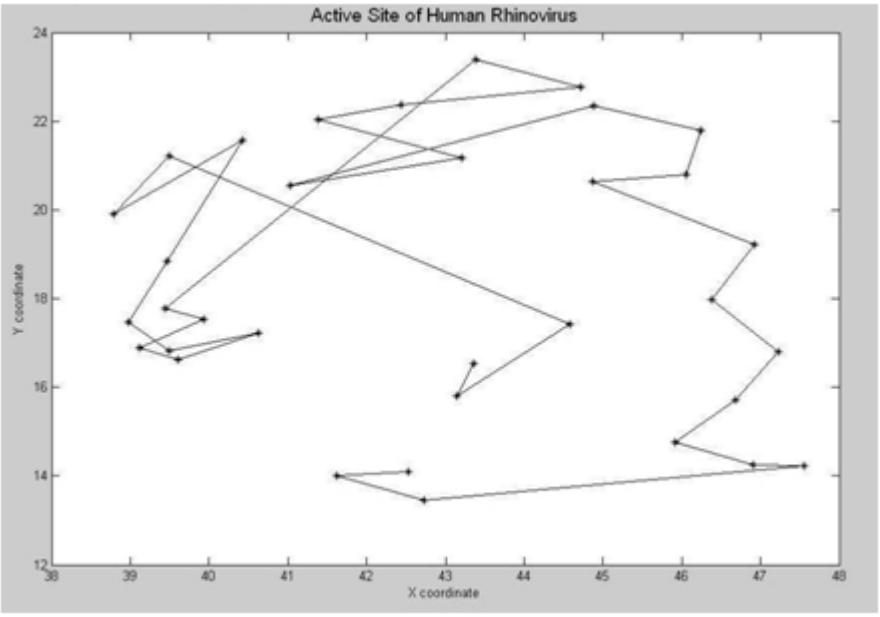


Figure 1. Active site of Human Rhinovirus stain 14

In the present context, we represent a ligand as a variable length tree structure. The length of the ligand will be determined according to the active site of the target protein as it cannot be determined before finding its structure; this can be seen in Figure 1, Figure 2 and Figure 3. We have considered the variable length tree having total 15 nodes. These nodes will be filled by at most 15 functional groups appropriately selected from the set of forty-five functional groups given in Figure 4.

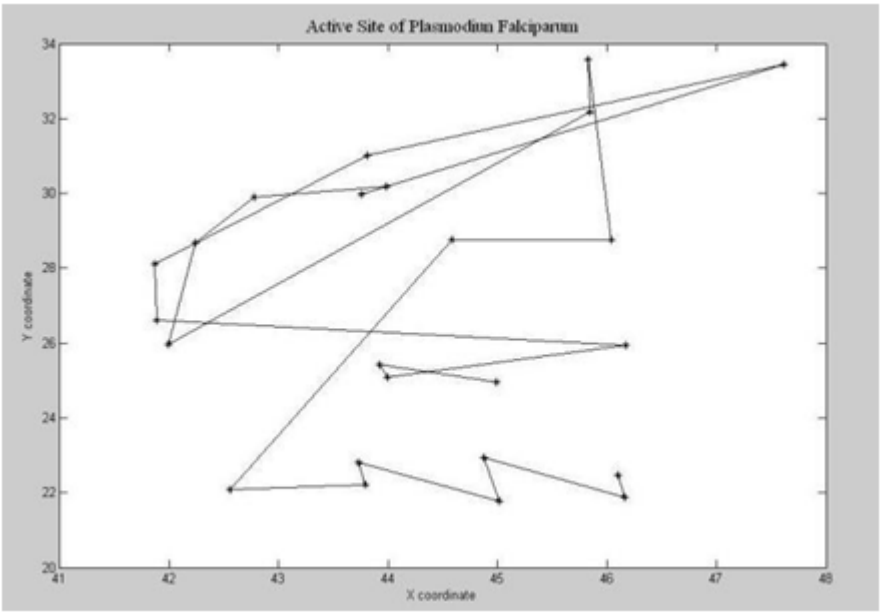


Figure 2. Active site of Plasmodium Falciparum

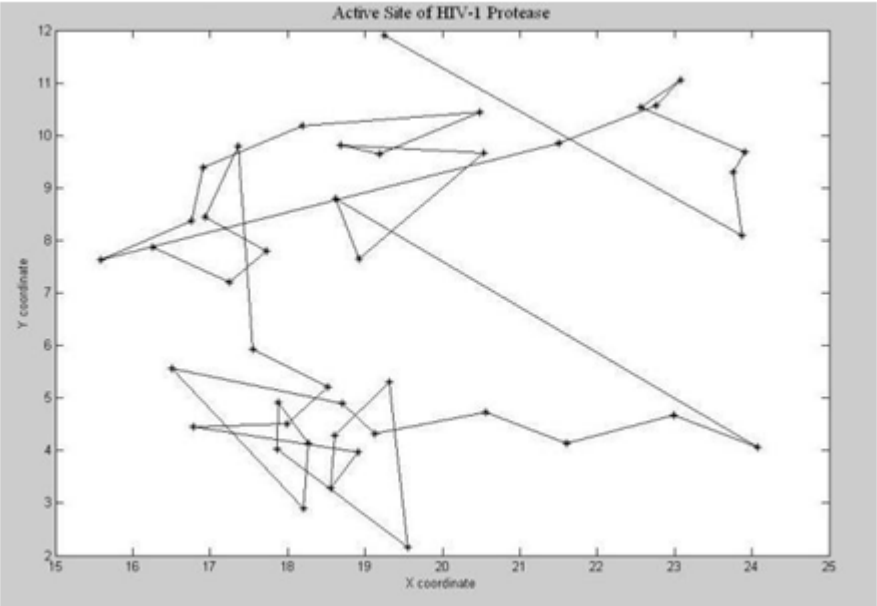


Figure 3. Active site of HIV-I protease

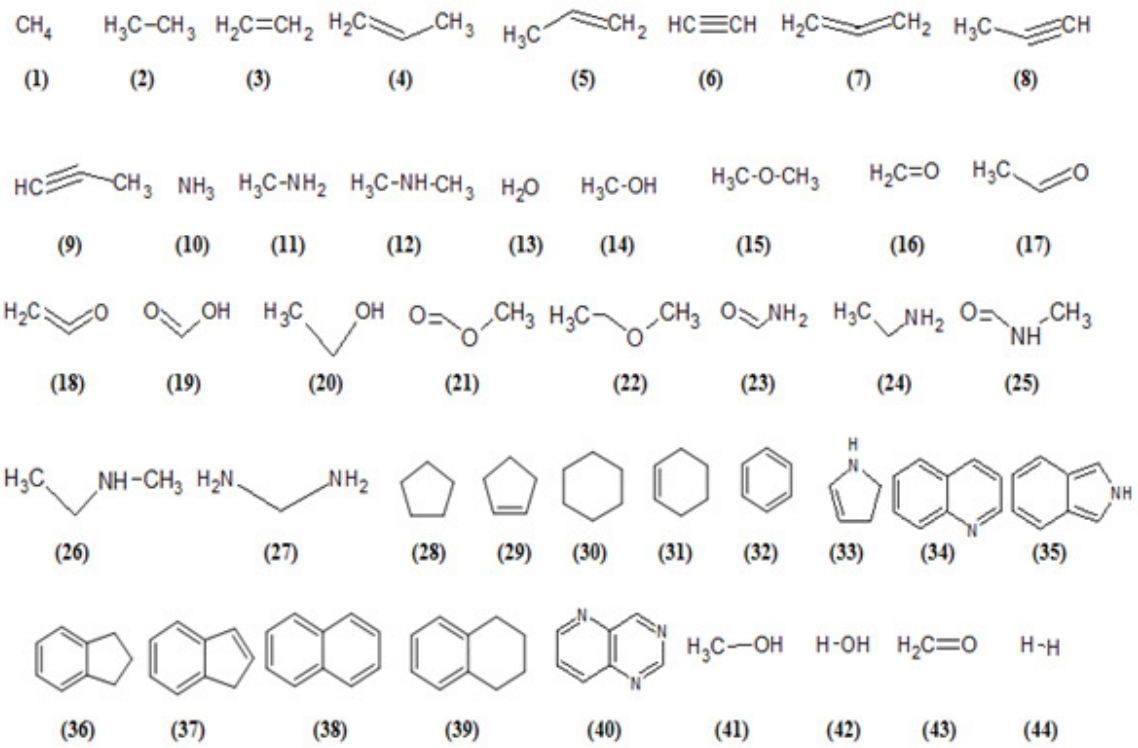


Figure 4. A set of forty five groups used to build ligands (45th group is the NULL group)

5. PARTICLE SWARM OPTIMIZATION ALGORITHM

PSO or Particle Swarm Optimization algorithm is an evolutionary computation technique that was developed by DT Eberhart and Dr. Kennedy [18]. This algorithm is inspired by social behaviour of bird flocking and fish schooling. It also begins with a random population and searches for optima by updating the population. In PSO, the potential solutions, called particles, flow through the whole search space by following the current optimum particles. Each particle has a speed vector and position vector to represent a possible solution.

PSO uses a simple rule. Before a change, each particle has three choices: (1) insists on oneself (2) moves towards the optimum it has met (3) moves towards the best the population has met. PSO reaches to a balanced position among these three choices. The algorithm is described as follows:

$$V_{i,d}^{k+1} = W \times V_{i,d}^{k} + C_1 \times R_1 \times (P_{i,d}^{pBest} - P_{i,d}^{k}) + C_2 \times R_2 \times (P_d^{gbest} - P_{i,d}^{k}) \quad (4)$$

$$P_{i,d}^{k+1} = P_{i,d}^{k} + V_{i,d}^{k+1} \quad (5)$$

Where V is the velocity, P is the position, k is number of iterations, d is number of particles, i is number of particle dimensions, W is inertia factor. C_1 and C_2 are acceleration coefficients. R_1 and R_2 are two other constants in the interval (0, 1).

6. SOLVING THE CONSTRAINT OPTIMIZATION PROBLEM USING PSO

Input:

1. Coordinates of active site.
2. Valency and length of 45 functional groups.

Output: Desired ligand structure L for receptor target protein P.

Begin
Call PSO (active_site_P);
End

Procedure PSO (active_site_P)

Begin

1. Initialize each particle Xi with 15 functional groups randomly chosen from the set of 45 groups at Figure 4.
2. Assign initial coordinates to each particle and call calculate energy (Xi) to find respective fitness.
3. Calculate fitness of each particle.
4. Calculate velocity and position of PSO parameters.
5. Update pBest and gBest as necessary.
6. Repeat steps 2 to 5 until a convergence criterion is satisfies or maximum iteration ends.

End

7. EXPERIMENTS AND RESULTS

The experiment was carried out in a simulated environment using MATLAB 2011. Population size for PSO is taken as 50 and the algorithm is run for 100 generation. In each generation, each of the particles is decoded to obtain the corresponding drug structure. Results are taken for different possible positions of the drug within the active site, and the evolved drug having the lowest energy value is taken as the solution. The two dimensional structure of the ligand is drawn using ChemSketch software [19].

We have applied our proposed PSO based approach on three different proteins viz. Rhino virus 14 Mutant N1105S (1RUC), Plasmodium Falciparum (TS-DHFR) (3DGA) and HIV-1 Protease (1W5X). Information about these three proteins is obtained from Protein Data Bank (http://www.rcsb.org/pdb/home/home.do) and the active site of a protein is obtained from http://dogsite.zbh.uni-hamburg.de/ . Table 1 shows the diseases caused by each of these proteins.

Table 1. Protein and corresponding disease caused by the protein.

Name of the protein	Disease caused by the protein
Rhino virus 14 Mutant N1105S (1RUC)	Cough and Cold
Plasmodium Falciparum (TS-DHFR) (3DGA)	Malaria
HIV-1 Protease (1W5X)	AIDS

We have compared our PSO based approach using Van Der Waals force with other available approaches. The docking energy values of the ligand–protein complexes are calculated for Human Rhinovirus and are given in Table 2. Lower inter-molecular energy values of the ligand imply better stability of the ligand. As seen in Table 2, our PSO based approach provides more stable ligands that are associated with lower energy values. Moreover, the length of the ligand is also checked with the size of active sites. The length of the ligand should always be less than the length of the active site. In our experiments, it was found that the length of the active site was $8.76\,\overset{0}{A}$ for Human Rhinovirus 14 Mutant N1105S, while the length of the ligand found using PSO was $8.72\,\overset{0}{A}$ which is less than the length of the active site. Hence, the ligand can fit quite well inside corresponding active site.

Table 2. Comparison of Result for Human Rhinovirus

Serial Number	Algorithm	Docking Energy (Kcal/mol)
1.	Fixed length GA	11.6454
2.	NBGA (fixed length)	11.5748
3.	VNBGA (Variable length)	8.1046
4.	Our Proposed Method (Variable length PSO)	6.5385

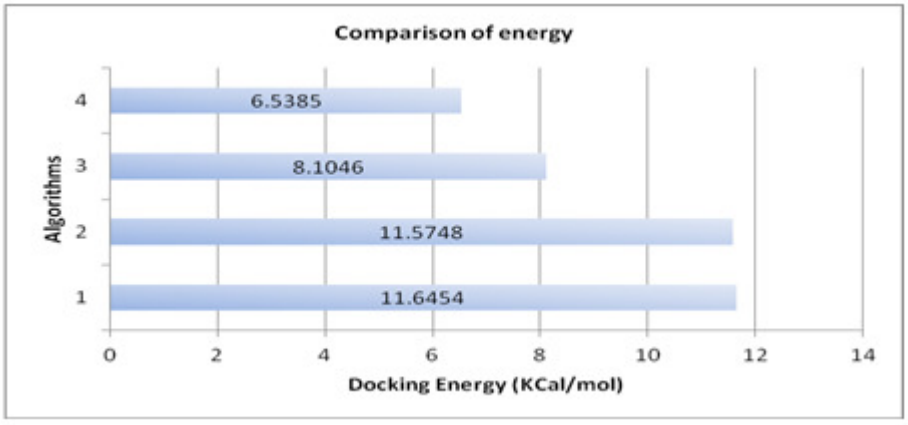


Figure 5. Comparison among algorithms mentioned in Table 2.

We have applied Electrostatics force of attraction along with Van Der Waal force to the Human Rhinovirus; the results obtained are 10.6352 Kcal/mol. However, Van Der Waal force is applied to Plasmodium Falciparum and HIV-1 Protease and the results obtained are 2.5477 Kcal/mol and 13.4303 Kcal/mol respectively. The application of Electrostatics force along with Van Der Waals force for Plasmodium Falciparum and HIV-1 Protease will be kept for future work.

The two dimensional structure of ligand molecule evolved using PSO are pictorially represented in Figure 6, 7 & 8 for Human Rhinovirus strain 14, Plasmodium Falciparum and HIV-1 Protease respectively. It is clear from the figures that the design molecules using PSO fill up the active sites quite well.

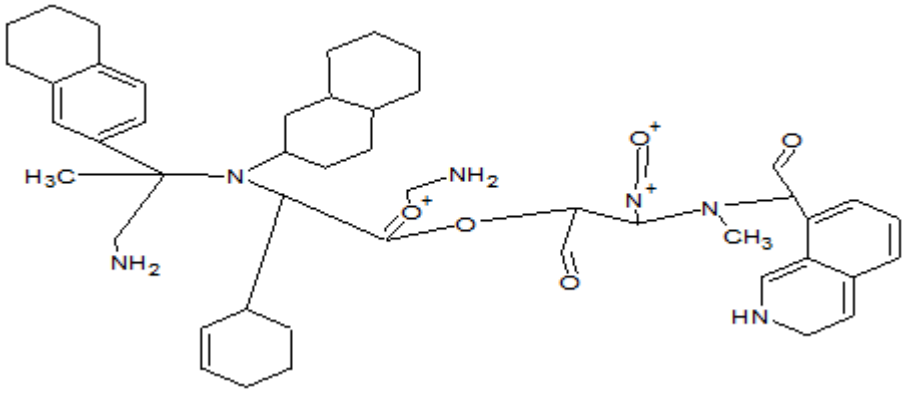


Figure 6. Generated Ligand Structure for Human Rhino Virus

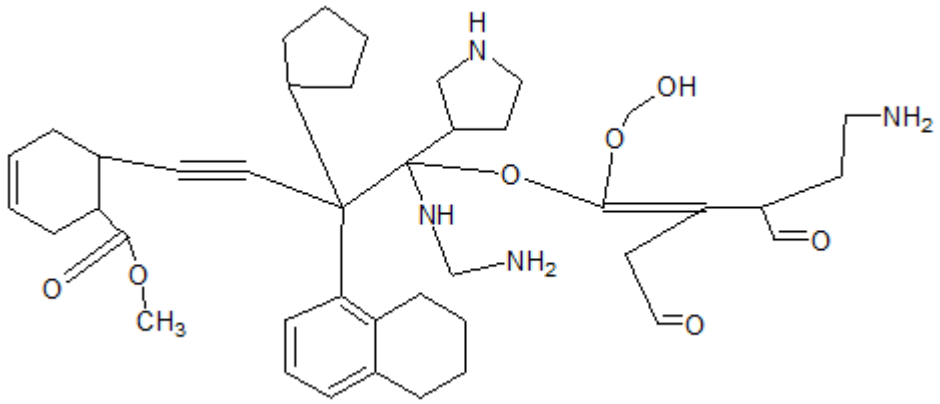


Figure 7. Generated Ligand Structure for Plasmodium Falciparum

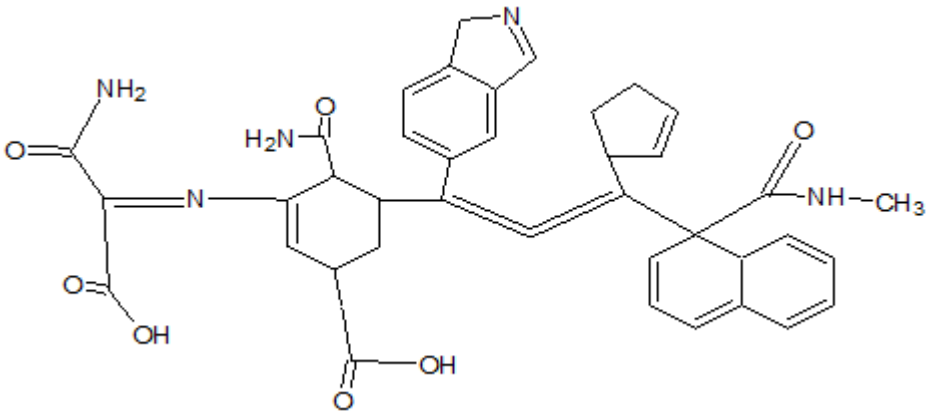


Figure 8. Generated Ligand Structure for HIV-I Protease

8. CONCLUSION AND FUTURE WORK

We can conclude that the proposed PSO based algorithm primarily optimizes protein-ligand inter-molecular docking energy. Our algorithm finds the ligand structure in such a way that each ligand can fit into corresponding active site very well. It also gives better result (less energy) than other methods. Moreover, using PSO, we can obtain more stable structure of ligand molecule. This proposed technique can be used to provide a powerful tool for the chemist to evolve molecular structure of ligand once the functional protein is given.

This work uses a two dimensional approach which has some limitations. A three dimensional approach using new data structure for the ligand will be our future research goal. Moreover, we have considered only inter-molecular forces but there are a number of intra-molecular forces like Bond stretching, angle binding, Dihedral angle. All these forces will give more accurate docking energy.

REFERENCES

[1] Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., ... & Bourne, P. E., (2000) "The protein data bank", Nucleic acids research,28(1), 235-242.

[2] Reyes, J. A., & Gilbert, D. (2007) "Prediction of protein-protein interactions using one-class classification methods and integrating diverse data".

[3] Chatterjee, P., Basu, S., Kundu, M., Nasipuri, M., & Plewczynski, D (2011) "PPI_SVM: Prediction of protein-protein interactions using machine learning, domain-domain affinities and frequency tables", Cellular & molecular biology letters, 16(2), 264-278.

[4] Moitessier, N., Englebienne, P., Lee, D., Lawandi, J., & Corbeil, A. C. (2008) "Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go", British journal of pharmacology,153(S1), S7-S26.

[5] Engelbrecht, A. P. (2005) "Fundamentals of computational swarm intelligence", (Vol. 1). Chichester: Wiley.

[6] Chen, H. M., Liu, B. F., Huang, H. L., Hwang, S. F., & Ho, S. Y. (2007) "SODOCK: swarm optimization for highly flexible protein–ligand docking", Journal of computational chemistry, 28(2), 612-623.

[7] Diwold, K., Himmelbach, D., Meier, R., Baldauf, C., & Middendorf, M. (2011) "Bonding as a swarm: applying bee nest-site selection behaviour to protein docking", In Proceedings of the 13th annual conference on Genetic and evolutionary computation (pp. 93-100). ACM.

[8] Eberhart, R. C., & Kennedy, J. (1995) "A new optimizer using particle swarm theory", In Proceedings of the sixth international symposium on micro machine and human science (Vol. 1, pp. 39-43).

[9] Andrew R. Leach. (2001) "Molecular modelling: principles and applications". Pearson Education.

[10] Goha, G. K. M., & Foster, J. A., (2000) "Evolving Molecules for Drug Design Using Genetic Algorithms", Proc. Int. Con. On Genetic and Evol. Computing, Morgan Kaufmann, 27-33.

[11] Ghosh, A., Ghosh, A., Chowdhury, A., & Konar, A. (2012) "An Evolutionary Approach to Drug-Design Using a Novel Neighbourhood Based Genetic Algorithm", arXiv preprint arXiv:1205.6412.

[12] Bandyopadhyay, S., Bagchi, A., & Maulik, U. (2005) "Active site driven ligand design: An evolutionary approach", Journal of bioinformatics and computational biology, 3(05), 1053-1070.

[13] Michalewicz, Z. (1996) "Genetic algorithms+ data structures= evolution programs", Springer Science & Business Media.

[14] Syswerda, G. (1991). "A study of reproduction in generational and steady state genetic algorithms", Foundations of genetic algorithms, 2, 94-101.

[15] Wei, D., Jiang, X., Zhou, L., Chen, J., Chen, Z., He, C., & Lai, L. (2008) "Discovery of multitarget inhibitors by combining molecular docking with common pharmacophore matching", Journal of medicinal chemistry, 51(24), 7882-7888.

[16] Namasivayam, V., & Günther, R. (2007) "PSO@AUTODOCK: A fast flexible molecular docking program based on swarm intelligence", Chemical biology & drug design, 70(6), 475-484.

[17] Guner, O. F. (2000) "Pharmacophore perception, development, and use in drug design", International University Line, La Jolla, CA, 29.

[18] Kennedy, J., & Eberhart, R. C. (1997) "A discrete binary version of the particle swarm algorithm", In Systems, Man, and Cybernetics, 1997. Computational Cybernetics and Simulation., 1997 IEEE International Conference on (Vol. 5, pp. 4104-4108). IEEE.

[19] Veit, R., & Gould, C. (2009) "Writing, reading, and research", Cengage Learning.

AUTHORS

Uttam K.Roy is presently Assistant Professor in the Department of Information Technology, Jadavpur University, Kolkata. He completed his M. Tech in Computer Science and Engineering, and PhD from Jadavpur University, Kolkata. For excellence in academics, he was awarded scholarships from UGC and Jadavpur University. In addition to his 12-year teaching experience, he has been a technical consultant and system administrator.

Dr Roy's research interests include bio-informatics, voice processing, optimization, and quantum computing. He is the sole author of two text books "Web Technologies" and "Advanced Java Programming", published by Oxford University Press, India in 2010. He also has contributed numerous research papers to various international journals, and has guided and supervised many postgraduate and Ph D dissertations.

Mainak Talukdar is presently working as a Programmer Analyst Trainee at Cognizant Technology Solutions, Kolkata. He has done Master of Engineering in Software Engineering from Jadavpur University. Earlier, he has done his Bachelor of Technology in Computer Science and Engineering from West Bengal University of Technology, Kolkata. He has about two year of industry experience.

Mainak's research interests include bio-informatics, data warehousing.

Anupam Ghosh is presently working as a Research Engineer at the Center for Indian Language Technology, Department of Computer Science and Engineering, Indian Institute of Technology Bombay, Mumbai. He completed his Master of Engineering from Jadavpur University, Kolkata. Earlier, He has done his Bachelor of Technology in Computer Science and Engineering from West Bengal University of Technology, Kolkata. He has Academic experience of about 2 years and Industry experience of about 2 year.

Anupam's research interests include bio-informatics, Natural Language Processing. He wishes to pursue his career in the field of Natural Language Process and Machine Leaning.

FAULT-TOLERANCE AWARE MULTI OBJECTIVE SCHEDULING ALGORITHM FOR TASK SCHEDULING IN COMPUTATIONAL GRID

Dinesh Prasad Sahu[1], Karan Singh[2] and Shiv Prakash[3]

[1,2]School of Computer and Systems Sciences,
Jawaharlal Nehru University, New Delhi-110067, India
[1]dinesh.sahu1230@gmail.com, [2]karan@jnu.ac.in
[3]Department of Chemical Engineering,
Indian Institute of Technology, Delhi-110 016, India
[3]shivprakash@chemical.iitd.ac.in

ABSTRACT

Computational Grid (CG) creates a large heterogeneous and distributed paradigm to manage and execute the applications which are computationally intensive. In grid scheduling tasks are assigned to the proper processors in the grid system to for its execution by considering the execution policy and the optimization objectives. In this paper, makespan and the fault-tolerance of the computational nodes of the grid which are the two important parameters for the task execution, are considered and tried to optimize it. As the grid scheduling is considered to be NP-Hard, so a meta-heuristics evolutionary based techniques are often used to find a solution for this. We have proposed a NSGA II for this purpose. The performance estimation of the proposed Fault tolerance Aware NSGA II (FTNSGA II) has been done by writing program in Matlab. The simulation results evaluates the performance of the all proposed algorithm and the results of proposed model is compared with existing model Min-Min and Max-Min algorithm which proves effectiveness of the model.

KEYWORDS

Computational Grid, Scheduling, NSGA II, Idle Time, MS, Fault-tolerance

1. INTRODUCTION

Primarily goal of computational grid (CG) is to fulfill the computational need of the tasks [1]. The computational grid [1] is combination of software and hardware infrastructure that facilitate right to use to enormous computational capabilities. The enormous computational capabilities are dependable, reliable, pervasive, and inexpensive even though the users and resources are geographically distributed. The fault tolerance is important parameter in terms of environment and cost. The tasks are migrated from idle nodes to busy nodes so that idle nodes operate in recoverable mode. There may be some nodes in grid which are idle and some which are busy.

David C. Wyld et al. (Eds) : ACITY, DPPR, VLSI, WiMNET, AIAA, CNDC - 2015
pp. 75–80, 2015. © CS & IT-CSCP 2015 DOI : 10.5121/csit.2015.51307

The challenging issue in computational grid is fault tolerance consumption due to cost and ecological problem [1]. The fault tolerance consumption based scheduling is key issue in this decade, by taking care of fault tolerance consumption. In order to decrease the fault tolerance consumption, the nodes in the grid have different computing power. Faster computing power of the nodes permits less time for task execution, but they consume much higher power. In fault tolerance -aware scheduling fault tolerance consuming at nodes is minimized during the task execution. In this paper, we explore Makespan (MS) and fault tolerance scheduling algorithm by using NSGA II for grid on a set of nodes. The organization of the work is as follows: The related work has been briefed in section 2 and the problem of the system has been described in section 3 and the proposed model is discussed in section 4. Experimental evaluation with respective observation is depicted in section 5 and finally conclusion is made in section 6.

2. RELATED WORK

Makespan, fault tolerance, turnaround time, availability, reliability etc. are some of the important parameters that are often optimized by scheduling the jobs appropriately on the grid nodes. The grid scheduling problem has been broadly deliberated in literature [1, 2, 7, 8]. Genetic Algorithm (GA) is highly used to find the solution of scheduling problem of CG, as the problem is NP-Hard [1, 17, 18, 19, 20, 21, 22]. Taking the effect of IPC into consideration, [9] uses GA for solving the problem of independent task scheduling in computational grid (CG). In [10], in place of a single, more than one task has been considered for allocation with load consideration using GA focusing on minimization of turnaround time in distributed computing systems. In [10], Load balancing, which is also to be taken care of and is an NP Hard problem, has been widely discussed in [26]. Load balancing on the grid nodes that uses GA has been elaborated in [11]. Both these paper considers load distribution and load variation among the nodes in CG. In [12], fault tolerance aware scheduling for independent tasks using GA is discussed. In this paper we modify the work of [12] by incorporating dependencies between tasks. The effectiveness of the model is studied by comparing with Min-Min, Max-Min. Another important parameter, security also find place in many papers. Security aware scheduling in computational grid (CG) is discussed in [22] with focus on security optimization. Parameter availability is discussed in [13] that demonstrate the availability metric [13], deals with the availability maximization for task scheduling problem of computational grid using GA.

3. THE PROPOSED MODEL

The proposed model, that considers MS optimization in computational grid, has been discussed as follows.

3.1 Fitness Function

Fitness function is derived using Queuing Theory [9, 26].

Then the MS can be compute by equation 3 as it is the maximum time taken for execution by the latest task.

$$MS = \max_{i=1}^{m}\left[\sum_{j=1}^{r}\left[\left(\frac{\lambda_i}{\mu_i(\mu_i-\lambda_i)}+\frac{1}{\mu_i}\right)\times\delta_{ji}\times NOI_j\right]\right] \text{----------(3)}$$

Consider a physical nodes of Computational Grid (CG), N is number of CG nodes. For computational simplicity, the nodes are assumed to be connected to each other following mess topology. The CG offers virtual connections various virtual are mapped to the CG. Let, The number of paths from S to Dusing the total number of path of equation (1), the maximization function for Fault Tolerance (FT) can be expressed as

$$Maximize\ Normalize\ FT = \max_{i=1,2,\ldots,N^p}(F_i^{path})/\sum_{i=1}^{n^p} F_i^{path} \quad (5)$$

where, F_i^{path}represents fault tolerance of i^{th}path from S to D.

The major components of the original NSGA-II are modified to adjust them into the framework of the optimization problem. The modified components are described below.

Minimize MS with MS≤ MUMS (6)

Maximize Normalize FT<=1 (7)

The proposed work is a multi-objective optimization problem where the required objectives are maximizing the fault tolerance and minimizing the MS achieved for tasks in CG. Multi-objective optimization is an important decision to find solutions of good quality. Since the conflicting complex nature of CG the multi-criteria makes optimization process complicated. An evolutionary algorithm is NSGA-II, which uses Pareto based evaluation with the capability to find several Pareto-optimal solutions in single iteration of simulation. Optimal fault tolerance of this paper is based on NSGA-II. Initial population is randomly generated and sorted according to fitness and the best half population is selected. The algorithm of NSGA-II is as follows [27]:

Step 1: A random population is initialized.
Step 2: Objective functions for all objectives and constraint are evaluated using equation 3 and 7.
Step 3: Front ranking of the population is done based on the dominance criteria.
Step 4: Crowding distance is calculated.
Step 5: Selection is performed using crowded binary tournament selection operator.
Step 6: Crossover and mutation operators are applied to generate an offspring population.
Step 7: Parent and offspring populations are combined and a non-dominated sorting is done.
Step 8: The parent population is replaced by the best members of the combined population.

In Step 3, each solution is assigned a non-domination rank (a smaller rank to a better non-dominated front). In Step 4, for each i^{th} solution of a particular front, density of solutions in its surrounding is estimated by taking average distance of two solutions on its either side along each of the objective [27]. This average distance is called the crowding distance. Selection is done based on the front rank of an individual and for solutions having same front rank, selection is done on the basis of their crowding distances (larger, the better)

3. EXPERIMENTAL EVALUATIONS AND RESULT

The proposed model is evaluated in this section by performing experimentations. The size of the input data set affects the convergence of the solution. The experiments are performed by writing the programs in Matlab version 9.0 and Java and integrated with Gridsim [4] to evaluate the performance on the system 500 GB secondary memory and 4 GB primary memory respectively. Simulation environment uses a random generator between the given ranges with uniform distribution. The following simulation parameters have been used in the experimentations. Input Parameter Values the Number of nodes 16-1000, Number of tasks 50-14000, Range of arrival Rate 1-100 MIPS, Range of task Size 2000-5000 MI and Range of processing speed 101-200 MIPS. Figure 4 contains these parameters and their values.

The performance evaluation is conducted for analysing FAGA (fault tolerance aware genetic algorithm) in the grid. The grid sizes are small (3 to 32 nodes and 10 to 512 tasks), medium (33 to 64 nodes and 513 to 1024 tasks), and large (65 to 128 nodes and 1025 to 2048 tasks. Figure 5 contains the type of grid and corresponding size in terms of the number of nodes and number of tasks. Normal distribution is used to randomly generate the capacity of the resources and the workload of tasks. It is expected that all submitted tasks to the system must be scheduled and all nodes of the system can be used. Above parameters are also used in [4, 12, 14].

For MS minimization minimize MS is taken as fitness which is discussed in [9]. For fault tolerance maximization we minimize TE is taken as fitness function. We run the simulation setup up-to 30 iterations and then average is taken for each result shown in this work.

Table 1. Makespan and tolerance by NSGA II

No. of nodes	No. of tasks	Makespan NSGA II	Makespan MIN-MIN	Makespan MAX-MIN	Tolerance by NSGA II
3	10	223.3	448.2955	448.2955	0.9981942
8	256	5578.8	10000000	3.5E+11	0.9826984
16	256	2674.2	6792966	6792966	0.9951092
32	256	10003.9	15782.03	15782.03	0.9991123
40	600	57297.3	2736774	2736774	0.9974828
80	1200	11123.5	2364124	2364124	0.9967923
50	700	50023.9	1086594	1086594	0.9769240
128	2048	153625.5	4105532	4105532	0.9860082

4. CONCLUSIONS

A huge number of developments has been introduced in Grid Computing in recent few years. Still, being an extremely heterogeneous system, grid poses a number of constraints. The research in the area of grid computing especially resource scheduling is going on. This work focuses on the proposal of scheduling in CG with emphasizing on the MS, and fault tolerance optimization using a NSGA II technique. NSGA II, a meta-heuristic, is a well-known procedure for solving complex multi-objective optimization problem. Performance of the proposed NSGA II method has been studied by carrying out the number of experiments and it is found that it performs well. Also, its comparative study with NSGA II model shows that it has an edge over Min-Min and Max-Min algorithms. The effectiveness of the model is also studies with scaled input and found that model is performing better than both Min-Min and Max-Min algorithms.

ACKNOWLEDGEMENTS

I would like to thank my guide Dr. Karan Singh for his kind suggestions and guidance. I like to thank Dr. Shiv Prakash for his help and guidance in this paper.

REFERENCES

[1] Foster and C. Kesselman, The Grid 2, Blueprint for a New Computing Infrastructure, Morgan Kaufmann Publishers is an Imprint of Elsevier, 2004.

[2] F. G. Berman, Fox Anthony and J.G. Hey, Grid Computing, Making the Global Infrastructure a Reality, John Wiley and Sons, 2003.

[3] M.R. Garey and D.S. Johnson, Computers and Intractability, A Guide to the Theory of NP-Completeness, W.H. Freeman and Company, New York, 1979.

[4] R. Buyya and Murshed M., "Gridsim a toolkit for the modeling and simulation of distributed resource management and scheduling for grid computing", Concurrency and Computation: Practice and Experience; vol. 14(13-15), pp.1175–1220, 2002.

[5] R. Buyya, D. Abramson and J. Giddy, An architecture for resource management and scheduling system in global computational grid. High Performance Computing Asia, China, IEEE CS Press: USA, vol. 1, pp. 283–289, 2000.

[6] Z. Shi and J. Dongarra, "Scheduling workflow applications on processors with different capabilities". Future Generation Computer Systems, vol. 22, pp. 665–675, 2006.

[7] S. Prakash and D.P. Vidyarthi, "A Novel Scheduling Model for Computational Grid using Quantum Genetic Algorithm", Journal of Supercomputing Springer, Vol. 65(2), pp.742-770, 2013.

[8] R. Buyya, D. Abramson and J. Giddy, "Nimrod/G: an architecture for resource management and scheduling system in a global computational grid", In :Proc. of the High Performance Computing Asia, China, IEEE, USA, vol. 1,pp. 283-289, 2000.

[9] S. Prakash and D. P. Vidyarthi, "Observations on Effect of IPC in GA Based Scheduling on Computational Grid". Int. J. of Grid and High Performance Computing vol. 4(1), pp. 67-80, 2012.

[10] S. Prakash and D.P. Vidyarthi, "Load Balancing in Computational Grid Using Genetic Algorithm", International Journal of Advances in Computing, Scientific and Academic Publishing, vol. 1(1), pp. 8-17, 2011.

[11] S. Prakash and D. P. Vidyarthi, "A model for load balancing in computational grid", In: Proc. of 18th annual international conference on High Performance Computing (HiPC11) Bangalore, Student Research Symposium , pp. 1-5, 2011.

[12] R. Kashyap and D.P. Vidyarthi, "Fault tolerance -aware scheduling model for computational grid". Concurrency and Computation: Practice and Experience, vol. 24(12), pp. 1377-1391, 2012.

[13] I. Koren and C.M. Krishna, Fault tolerant systems. Morgan Kaufmann is an imprint of Elsevier, 2007.

[14] S. Prakash and D. P. Vidyarthi, "Maximizing availability for task scheduling in computational grid using GA", Concurrency and Computation: Practice and Experience, Wiley, vol. 27(1), pp. 193–210, 2015.

[15] T.D. Braun, H.J. Sigel, and N. Beck, "A comparison of eleven static heuristic for mapping a class of independent tasks onto heterogeneous distributed computing systems", Journal of Parallel and Distributed Computing, vol. 61, pp. 810–837, 2001.

[16] J.H. Abawajy, "Automatic Job Scheduling Policy for Grid Computing", LNCS, Springer-Verlag Berlin Heidelberg, vol. 3516, pp. 101-104, 2005.

[17] F. Xhafa and A. Abraham, "Meta-heuristics for Grid Scheduling Problems", Studies in Computational Intelligence Series, Springer, vol. 146, pp. 1–37, 2008.

[18] F. Xhafa and A. Abraham, "Computational models and heuristic methods for Grid Scheduling problems", Future Generation Computer Systems, Elsevier, vol. 26, pp. 608-621, 2010.

[19] F. Xhafa and A. Abraham, "A Genetic Algorithm Based Schedulers for Grid Computing Systems" International Journal of Innovative Computing, Information and Control, vol. 3(6), pp.1–19, 2007.

[20] W. N. Chen and J. Zhang, "An ant colony optimization approach to grid work flow scheduling problem with various QoS requirements", IEEE Transactions on Systems, Man and Cybernetics--Part C: Applications and Reviews, vol. 39(1), pp. 29-43, 2009.

[21] S. Prakash and D. P. Vidyarthi "A Hybrid GABFO Approach for Scheduling in Computational Grid", International Journal of Applied Evolutionary Computation (IJAEC) vol. 5(3), pp. 57-83, 2014.
[22] D.A. Menasce and E. Casalicchiop, Quality of service aspects and metric in grid computing. In: Proc. of Computer Measurement Group Conference, pp. 511-532, 2004.
[23] A. Rajni and I. Chana, "Formal QoS policy based grid resource provisioning framework", J. Grid Computing, vol. 10(2), pp. 249–264, 2012.
[24] R.B. Cooper, Introduction to Queuing Theory, 2nd Edition, Elsevier North Holland Publications, 1981.
[25] S.Prakash and D.P. Vidyarthi, "Immune Genetic Algorithm for Scheduling in Computational Grid", Journal of Bio-Inspired Computing, vol. 6(6), pp. 397-408, 2014.
[26] C. Kumar, S. Prakash, T. Kumar and D. P. Sahu, "Variant of genetic algorithm and its applications", International Journal of Artificial Intelligence and Neural MCGs, vol. 4(4), pp. 8-12, 2014.
[27] K. Deb, S. Agrawal, A. Pratap, and T. Meyarivan. A fast and elitist multi-objective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation, 6(2):182–197, 2002.
[28] S. Prakash, D.P. Sahu, Karan Singh, "Maximizing Availability and Minimizing Markesan for Task Scheduling in Grid Computing using NSGA II", 2nd international conference for computer and communication technology, LNCS, Springer, 1-6, 2015.
[29] D. P. Sahu, K. Singh and S. Prakash"A Review on Resource Scheduling Models to Optimize Quality of Service Parameters in Grid Computing using Meta-heuristics", International Journal of Computer Applications 114(8):pp.1-4, 2015.

AUTHORS

Dinesh Prasad Sahu received the Master degree (Computer Science & Application) M.Tech (Computer Science & Application) from Jawaharlal Nehru University, New Delhi, India. Currently, He is doing Ph.D. (Computer Science & Engineering) under the guidance of Dr. Karan Singh, from Jawaharlal Nehru University, New Delhi, India. Currently, he is working in school of Computer & Systems Sciences, Jawaharlal Nehru University, New Delhi. His primary research interests include parallel and distributed system and Grid Computing. He has published 3 papers in proceedings of peer-reviewed Conferences.

Karan Singh (M-1981) received the Engineering degree (Computer Science & Engineering) from Kamala Nehru Institute of Technology, Sultanpur, UP, India and the M.Tech (Computer Science & Engineering) from Motilal Nehru National Institute of Technology UP, India. He is Ph.D. (Computer Science & Engineering) from MNNIT Allahabad deemed university. Currently, he is working in school of Computer & Systems Sciences, JawaharlalNehru University, New Delhi. His primary research interests are in computer network and computer network security. He is reviewer of IEEE & Elsevier conferences and reviewer of International journals. I have research papers (8 Journal, 9 IEEE conferences, 2 Elsevier Conference, 15 National and International conference) and 2 are accepted. He2 is organizer of various workshop, Conference and training. Recently, Dr. Karan work as General Chair of Qshine 2013. Dr. Singh has been joined as Professional member of IEEE, ACM, CSTA, CSI, IACSIT, ICST, IAENG, ACEEE, ISOC and IEEE computer society.

Shiv Prakash received his M.Tech and PhD in Computer Science from School of Computer and System Sciences, Jawaharlal Nehru University, New Delhi, India in 2010 and 2014. He has published around 15 papers in various international journals and around 5 papers in proceedings of peer-reviewed conferences. He is a member of the IEEE and ACM. His research interest includes parallel/distributed system and grid computing.

HASHTAG RECOMMENDATION SYSTEM IN A P2P SOCIAL NETWORKING APPLICATION

Keerthi Nelaturu[1], Ying Qiao[1], Iluju Kiringa[1] and TetHin Yeap[1]

[1]School of Electrical Engineering and Computer Science,
University of Ottawa, ON, Canada
{knela006, yqiao074, iluju.kiringa, tyeap}@uottawa.ca

ABSTRACT

In this paper focus is on developing a hashtag recommendation system for an online social network application with a Peer-to-Peer infrastructure motivated by BestPeer++ architecture and BATON overlay structure. A user may invoke a recommendation procedure while writing the content. After being invoked, the recommendation procedure returns a list of candidate hashtags, and the user may select one hashtag from the list and embed it into the content. The proposed approach uses Latent Dirichlet Allocation (LDA) topic model to derive the latent or hidden topics of different content. LDA topic model is a well-developed data mining algorithm and generally effective in analysing text documents with different lengths. The topic model is used to identify the candidate hashtags that are associated with the texts in the published content through their association with the derived hidden topics.

The experiments for evaluating the recommendation approach were fed with the tweets published in Twitter. Hit-rate of recommendation is considered as an evaluation metricfor our experiments. Hit-rate is the percentage of the selected or relevant hashtags contained in candidate hashtags. Our experiment results show that the hit-rate above 50% is observed when we use a method of recommendation approach independently. Also, for the case that both similar user and user preferences are considered at the same time, the hit-rate improved to 87% and 92% for top-5 and top-10 candidate recommendations respectively.

KEYWORDS

Bestpeer, baton, Hashtag, topic model, hit-rate and peer-to-peer networks.

1. INTRODUCTION

Most of the current social networks adopt centralized server architecture. This kind of architecture has both its pros and cons. In centralized architecture, we have all the applications running with their data at one location, at which one or more large computers are connected. Pros include ease of maintenance, any administration or upgrade on the system can be easily done across the components of all the applications. Backup and restore mechanisms are easy to implement since its just one central location and security mechanisms can be incorporated in a simple manner. On the other hand cons include, bottleneck in performance and privacy concerns

David C. Wyld et al. (Eds) : ACITY, DPPR, VLSI, WiMNET, AIAA, CNDC - 2015
pp. 81–93, 2015. © CS & IT-CSCP 2015 DOI : 10.5121/csit.2015.51308

from user data perspective. In order to avoid these defects a different line of architecture pattern called the distributed or peer-to-peer architectures are being employed. Peer-to-peer (P2P) systems support for user data privacy, scalability, and availability avoiding single point of failure. Keeping this in view, we are working towards the development of a unique social networking application, which has peer-to-peer architecture. The architecture is inspired from BestPeer++[4] and BATON overlay network [8]. In any social networking application as the user-generated content increases it becomes hard to organize ones own data. Tagging has been a way of organizing data in many of the social networking sites like Facebook[1] and Twitter[2]. We make use of Hashtags, which is one way to tag content. Hashtags are short words with continuous characters without any space in between. They are identified by the presence of '#' before the words. They can be used anywhere within the messages, phrases etc. They have been mainly used for categorizing or highlighting an event, topic, news, individuals etc [15]. This concept has been employed in many social networking sites till date and has become popular with the start of Twitter Social Networking website. Until now these have been used for media broadcasting and business, promotions etc [13]. We developed a hash tag recommendation approach for our online social networking platform to suggest suitable hash tags to a user.

The paper is organized as follows. Section 2 covers the background and related work, which includes details on popular recommendation system techniques.Section 3, illustrates the architecture of our peer-to-peer based social networking application and its high level components. In this section, we discuss the modified implementations of BestPeer++ architecture and BATONoverlay network. Section 4, we discuss in detail the hashtag recommendation methodology.Section 5, presents the details on the datasets used for experiments, the test setup environment and all of the experiments performed to ensure the correctness of algorithms and to calculate the performance of the algorithms. Section 6, we present the conclusion and future work.

2. BACKGROUND AND RELATED WORK

2.1. BestPeer++ Architecture

BestPeer++[4] is a cloud service model. Any business that wants to use the service just has to register themselves and create a BestPeer++ instance, into which they can export data for further processing. This also gives an option for pay-as-you-go query processing model with the help of cloud computing. There are two main components in BestPeer++ - core and adapter. Adapter has two parts, one is an interface to the service and the other part contains adapters, which implement this interface with the help of service provider APIs. The core component consists of query processing and the P2P overlay for serving responses to the queries. There are two kinds of elements in core, bootstrap peer and normal peer. When a business creates an instance, a database server is assigned to that particular instance. This server is then included into the structured P2P overlay arrangement, along with all the other servers. So a normal peer here is a server of a particular business instance. Figure 1 shows the components of the BestPeer++ architecture.

Responsibilities are divided between bootstrap peer and normal peer. The whole network has a single bootstrap peer. This is the server through which normal peers try to join the network. It

[1]Facebook, https://Www.facebook.com/.
[2]*Twitter.* (2015). https://about.twitter.com/company

works like an administrator for the network. Some of the tasks performed by this bootstrap peer are - auto scaling (when an instance exceeds its storage or to perform load balancing), auto fail-over (when a node in the P2P overlay has failed and had to be removed from the network) and the main task of node joining/leaving. For a normal peer, primary effort goes in data loading and indexing. It also does the schema mapping, query processing and execution, along with data loading. When a new business is added to the network, data is loaded from the corporate production to the instance. When this process is being done, normal peer tries to do schema mapping i.e mapping the local business schema to the global peer schema.All the normal peers are organized in P2P overlay called, BATON(Balanced tree overlay network) [8]. This is the crux for BestPeer++ functionality. It provides the interface for node joining, leaving, adding or removing data etc. It arranges all nodes in tree structure. BATON allows for processing both exact and range queries. BATON also provides for three types of indexes - table, column and range. BestPeer++ also provides for role-based distributed access control. When a node fails, all the queries are held up until the backup is restored on to the system. With all these features, cloud computing, database and P2P overlay support BestPeer++ is highlighted as a better data sharing application than any other P2P data sharing systems available. Hence, we choose the same for our P2P social networking application.

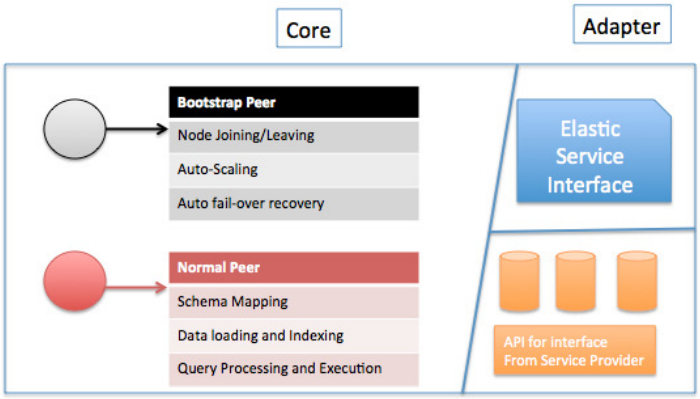


Figure 1: BestPeer++ Architecture components

2.2. Recommendation Systems

Recommendation system is a facility that has been used in web applications for “predicting the user responses to options”[11]. It involves the technique, which is used to make suggestions to the users based on certain selected criterion. Recommendation methods have been classified into two major types: Content-based and Collaborative Filtering. They focus on different perspectives while making recommendations. Content-based methods are specifically on the “properties of items”, and collaborative filtering methods are on the “relationship between users and items”[11]. Content-based recommendation systems (CBR), base their suggestions on the similarity between the items. One major reason behind this content-based approach is that a user always selects similar items. Therefore, during the recommending operation, these systems match the profiles of the items with the profiles of the users. Content-based recommendation becomes better approach for suggestions in a system where there are not many users. However, these systems also have some corresponding limitations [12]. First, the annotations that are added to the content either automatically or manually always will have limited details. It has been seen that the keywords identified for web pages might not contain any information about the media embedded in these web pages. Another limitation is termed as over-specialization. When recommendation is based

on the content already rated by a user, the concentration is restricted to the area already visited by the user. The data outside the domain of the likes of the user might not be considered. Although the dependency with other users in the system is reduced, for a new user, proper recommendations cannot be made until sufficient data about the user's interests have been collected.

Collaborative-filtering recommendation systems, base their suggestions on the similarity of the user's choices on two items. For example in [5], Collaborative-filtering (CF) method employs the nearest-neighbour algorithms to recommend products to a target customer based on the preferences of the neighbours, who have similar interests as of this customer. Though CF methods avoid some of the limitations of the CBR methods mentioned above, there are some drawbacks even with the CF methods. One of them is the same as CBR methods. In order to compare, the interests of a new user with those of others, the CF methods need the information about the ratings or items the user is interested is needed. Other problems are related to the new content added to the application and also sparsity issues. For a new item, it takes some substantial amount of time for the system to collect rating details from other users. Some content might be rated high by a small number of users with peculiar interests. Considering user's profile information apart from the rating data will avoid such scarcity issues.

To avoid limitations of different techniques, many applications implement hybrid recommendation strategies wherein they use both content-based and collaborative filtering techniques are adopted. For our hashtag recommendation, we use hybrid recommendation technique.

2.3. Popular recommendation system techniques

Term Frequency-Inverse Document Frequency (TF-IDF) is the method generally used in text mining and information retrieval systems [3]. It calculates the importance of a word in a document against a corpus. TF-IDF also is offset by the frequency of the word in the corpus. A vector space model indicates the weight in this kind of measure. Term Frequency (TF) of a term in a particular document measures the number of times a term appears. Inverse document frequency (IDF) of a term is calculated upon overall corpus not just one document. It gives the importance of a term in the complete document corpus. Below equation shows the TF-IDF weight calculation, where $m_{ij}$is the number of times a term (t_i) appears in document (d_j), m_i is the number of documents the term has appeared in. M is the total number of documents in the whole corpus[14].

$$tf_{ij} = \frac{m_{ij}}{\sum_k m_{kj}} \qquad idf_i = log\frac{M}{m_i} \qquad TF - IDF_{ij} = tf_{ij}.idf_i$$

In this method, generally document length also plays as a factor. Longer documents tend to have higher values due to the increased number of words and word repetitions. Hence, while calculating the weights of the terms, this approach always normalizes these weights with the length of the documents.

The other technique generally used by recommendation systems is topic models. Topic models are based on the idea that documents are the mixture of topics, where a topic is a probability

distribution over words[17]. A topics model is a generative model. In a generative model, a joint probability distribution is defined over a set of observed and hidden random variables. The joint distribution can be used to generate observable random variables in a generative process. Furthermore, a conditional distribution on hidden random variables can be obtained with the use of the joint distribution and the observed variables. The conditional distribution is also termed as posterior distribution[2]. A topic model always revolves around word and document distributions progressively.

Latent Dirichlet Allocation (LDA) is the one of the simplest topic models. The intuition for LDA is the same as all the other topic models. But the main characteristic is that, all the documents in LDA share the same set of topics. Each document has a probability over each of these topics. The computational problem for LDA is to observe a set of documents and identify the topic-document and topic-word distributions. These probability distributions can further be used for inferring the topic structure of any other documents. LDA also follows the generative model definition. In LDA, the observed variables would be the words of the documents, and the hidden random variables would be the topics.

Here, we describe LDA more formally as defining the topic mixture for each document i.eP(t/d), with a topic mentioned by a distribution over words i.e $P(w_i/t)$ as shown in below equation, where $P(w_i/d)$ is the probability of ith word in a given document d and $t_i$is the topic and $P(t_i = j/d)$ is the probability of identifying a word(w_i) from topic j appearing in document d. $P(w_i/t_i = j)$ is the probability of picking a word from a topic j.

$$P(w_i/d) = \sum_{j=1}^{t} P(w_i/t_i = j)P(t_i = j/d)$$

The topic-document P(t/d) and topic-word P(wi/t) distributions can be estimated by using a corpus of documents[10]. In general convention, θ denotes the topic distributions and φ denotes topic-word distributions. Gibbs sampling algorithm is one of the approaches used for extracting topics from a corpus. It uses an iterative process, which stops until the target distribution is achieved. In an iterative round, each word in the corpus is considered and the estimations for the probability of assigning that word to a topic is done with below equation, conditioned on other word tokens in the same topic. From this conditioned distribution, a topic is sampled and stored as a new topic assignment[17].

$$P(t_i = j|t_{-i}, w_i, d_i, .)\alpha \frac{C_{w,j}^{WT} + \beta}{\sum_{w=1}^{W} C_{wj}^{WT} + W\beta} \frac{C_{d,j}^{DT} + \alpha}{\sum_{t=1}^{T} C_{d,t}^{DT} + T\alpha}$$

In the equation, C^{WT} maintains count of all topic-word assignments, C^{DT} has the document- topic assignments, t_{-i} represents all topic-term and document-topic assignments except for the current assignment t_i, for word w_i, αand β are the hyper parameters for the Dirichlet priors, works as smoothing factor for the counts. The estimated distributions can be further used in the operations of a recommendation system.

2.4. Related Work

Most of the recommendation proposals ([20], [9], [6], [19]) use only content-based methodology. Using both content-based and collaborative filtering techniques like our proposal Jieying She and Lei Chen's approach [16] results in better hit-rate. Godin et al[6] even though has good hit-rate results, suggests only keywords for the recommendations. The authors do not consider the hash tags already in use and also no collaborative filtering techniques implemented, which reduces the scope of hash tags considered. As per our knowledge, none of the recommendations developed are for Peer-to-Peer network topology. By using the P2P features like scalability and maintenance our approach could achieve a better performance over the other studies.

3. ARCHITECTURE AND COMPONENTS

BestPeer++ is a two-layered architecture. In the current P2P application we have three-tier architecture - bootstrap peer, server peer and client peer as in Figure 2.

3.1 Client Peer

In our application, each user whoever wants to join the network need to use a client side user interface on their PC or mobile device. This user is called the Client Peer. We do not store any data on the client side. All of the data pertaining to a user is stored in the database on the server side. The User Interface helps the user in interacting with the application.

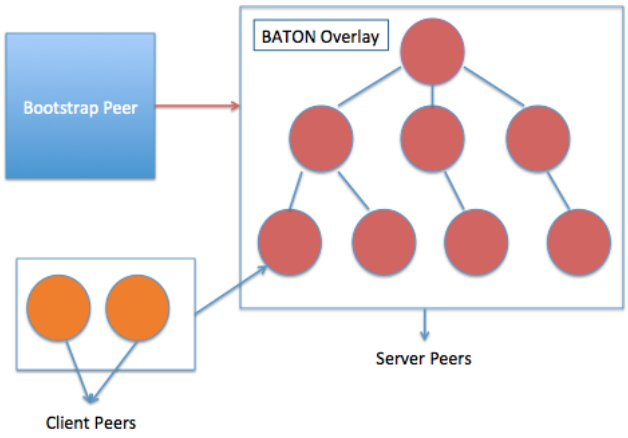


Figure 2: Our P2P Architecture

3.2 Bootstrap Peer

Bootstrap peer in the current architecture, has the same administrator role as in BestPeer++. Single bootstrap peer node accounts for the health of the whole network. Monitors the node joining and leaving. The auto-failover and auto-scaling supposed to be in the bootstrap peer have not been implemented in our system. Users need to register and login via the bootstrap peer each time they connect to the network. Apart from the components in BestPeer++, we also store the user profile and friendships information in Bootstrap peer.

3.3 Server Peer

Client data is stored in the Server Peer nodes. Each server peer is responsible for more than one client node at a time. All of the server peer nodes form the BATON overlay structure. User login information is stored even in the server peers. Major concerns for the server peers lies in the management of P2P overlay structure and user data. Each server is responsible for clients within a particular URI (User Resource Index) range.

Server Peer is the alias for Normal Peer in BestPeer++. Most of the functionalities in normal peer have been imported to server peers with some updates. The schema-mapping module was discontinued as all the data exchanged in the application has the same mapping. Data loader is used during the data retrieval process. Data Indexer is major for the BATON overlay network, as each of the data stored in server depends on the range. It also helps during the forward and lookup requests. For the query execution, we used JPA instead of pure SQL language [18]. BATON tree node information is also stored with the server along with the physical details of bootstrap peer.

4. HASHTAG RECOMMENDATION APPROACH

4.1. Proposed Approach

The proposed hashtag recommendation approach lists out hashtag candidates for a content entered by the user. If no related hashtags are found, this approach may suggest the user with the hashtags that have been used previously or with those related to the user or to the content. The approach also advises the user with some keywords for creating a new hashtag. We adopt a hybrid recommendation system for our social network platform considering both types of recommendation: content and collaborative filtering approaches. Most of the hashtag recommendation systems have lagged in two issues. First, they use only one of the recommendation approaches. In the case that an approach is chosen, a major part of hashtags the user might be interested in is being omitted. For example, the content-based techniques might not include some of the tags being created by similar users or the friends in the suggested tags. Similarly, the collaborative filtering based techniques might neglect the tags related to the posted content or those popular in the overall system. Second, as per our knowledge, none of the ideas reviewed till now have given a user an opportunity to choose the recommendation method he might be interested in.

Hence, considering these drawbacks in the previous research, our approach contains several recommendation modes. The users may control the recommendation system by selecting one or multiple modes. They receive the candidate hashtags recommended by the selected modes. These modes are classified into the following categories. The categories considered are:

1. Global content common for all of the users
2. User preferences evaluated based on their content previously added
3. Hashtags created by users with similar preferences as current user
4. Hashtags created by the friends of the user and are related to the users content being created
5. Overall popular hashtags in the whole social network platform

Also, in the case that any of the methods returns zero tags, the proposed approach even recommends with keywords based on the chosen mode.

4.2. Implementation

Unlike twitter, which has restricted the lengths of the text for its tweets, content with different types in our application can have varied lengths without any restrictions. In such cases, topic model for recommendation systems is a better technique. For our approach we considered to adopt topic analysis technique to evaluate the content similarities and user preferences on content. We further chose Latent Dirichlet Allocation (LDA) process using Gibbs Sampling method for topicclassification. Most of the research in topic models considers only topic-word distributions. The proposed approach goes further to the next step and extracts "topic-hashtag" probability distributions. The LDA process is done in three phases as shown in Figure 3.

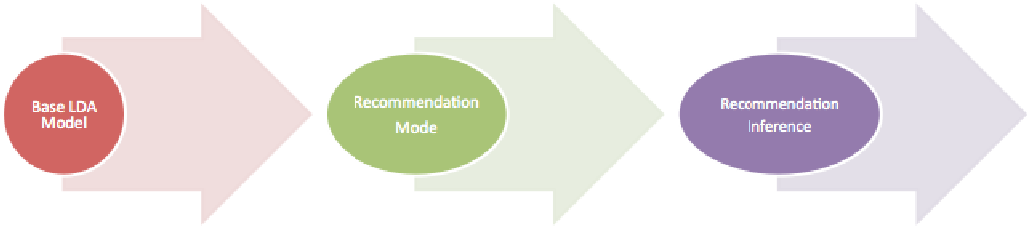


Figure 3: Steps in our LDA processing

Training or Base LDA Model: In the training phase, random content collected across various networks is passed to the procedure implementing the Gibbs algorithm. The procedure estimates the initial topic-word and document distributions of the Base LDA model. The documents in this situation refer to any single post, comment or article.

Estimation or Recommendation Mode: This is phase at which content passed for LDA procedure differs from each of the categories mentioned before. For each mode selected, we pass in the content and generate the updated topic- word, document distributions along with topic-hashtag distribution. For each of the topics and documents, we calculate hashtag distribution over documents can be calculated using below equation where d is the document, h is the hashtag and t is the topic.

$$P(h/d) = \sum_{i=1}^{n} P(h/t_i)\, P(t_i/d)$$

Recommendation Inference: This is the phase in which a user is suggested with candidate hashtags. Content a user enters is passed to the model to evaluate the topic-hashtag distributions and order them according to their probability scores. This procedure also uses the equation above only d refers to the union of the words in the content.

5. EXPERIMENTS AND RESULTS

In this section, we discuss the experiments performed for verifying the correctness of the hashtag algorithms and evaluate the effectiveness of these algorithms.

5.1. Experimental Setup

All of the experiments were performed on a standalone computer with 16GB RAM and MAC OS X or Windows 7 operating systems. We developed a prototype application with JUnit test cases for hashtag implementation. In our prototype, we have two server peers, one bootstrap peer and fifteen client peers. We use JGibbLDA library[3] for performing LDA operations. We set the default values for hyper-parameters $\alpha = 50.0/k$ and $\beta = 0.1$ as suggested in [7] by Griffiths and Steyvers where k is the number of topics considered. For all the experiments we run the LDA operations through 500 iterations of Gibbs Sampling. The contents used by these experiments are, datasets obtained from three different sources. The first was from Textual Retrieval Conference (TREC) 2011 micro blog track[4]. We choose Tweets2011 corpus. This corpus comprises of 16 million tweets collected over a period of two weeks between 24th January 2011 until 8th February 2011. We used Twitter tools API provided by TREC Microblog track to extract tweets. The second source was Twitter web site. We use the Twitter Streaming API to extract tweets. We captured 10000 tweets with trending topics in specific intervals of time for two days. Third source is from Sentiment 140[5]project created by the students from Stanford University for the purpose of Sentiment analysis of topics in tweets. This collection consisted of two datasets: one was training data with 1,60,0000 tweets and one was test data with 500 tweets. Sentiment140 data is pre-processed, where any special characters or emoticons are removed. Before passing the data to the LDA functions, we selected the tweets in these datasets, removed the special characters or any characters other than English letters. The special characters "#" are kept, since it indicating the beginning of a hashtag.

5.2. Experiments and Results

We perform experiments on each of the recommendation modes mentioned in our proposed approach. For evaluating the effectiveness of this recommendation approach, we consider hit-rate of the results from an execution of a recommendation activity as the criteria. A recommendation activity starts from invoking the recommendation function on content upon a user's request to returning the results to the user. The equation to calculate the Hit-rate of the results is defined below. We identify a result as hit if atleast one of the recommended hashtags is a hashtag used for the content.

$$Hit-rate = \frac{Number\ of\ hits}{Number\ of\ content\ items\ considered}$$

There were three sets of experiments performed. For all of the experiments, apart from comparing the actual hashtags used in the content, we also performed subjective evaluation with the five evaluators. The evaluators where asked to mark the recommended hashtags as relevant and non-relevant. Majority views of the votes were considered for the final results.

First experiment compares the hit rate percentage over the number of topics for each of the recommendation modes except for the mode in which we recommend hashtags based on their overall popularity. Figure 4 shows the graph plotted for four categories of recommendations with hit-rate against topics. Initially for all the methods we started of with 50 topics for the LDA.

[3]JGibbLDA, http://Jgibblda.sourceforge.net/.
[4]TREC Twitter2011 datasets, http://Trec.nist.gov/data/tweets/.
[5]Sentiment 140, http://Www.sentiment140.com/.

Since topic-hashtag distribution is the main case that we consider for our proposal, with more topics we expected more hashtags.

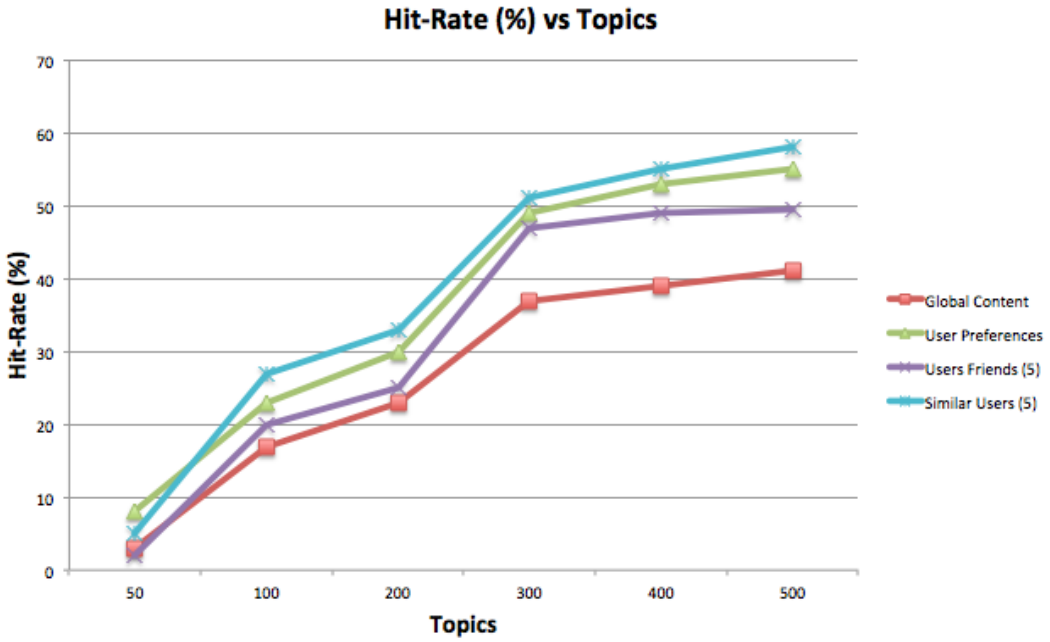


Figure 4: Hit-rate Vs Number of Topics

Recommendation with global content was not satisfactory, as it wouldn't consider any of the user content or user topics of interest. Maximum was 41.3% of hit-rate with global content that too at 500 topics. The results for User Preference based and similar users based recommendations were promising and we were able to see 55% and 57.6% of hit-rate respectively. For the recommendations from similar user and friends we used 5 clients as the users under comparison. Recommendation mode using friends content and interest could give approximately 50% of hit-rate. The other observation we made was on the cases with the number of topics between 300-500 topics; however there was not much of improvement with the results. So, for our application 300 topics would be the ideal number of topics to be considered. Overall Popularity based recommendation needs more server nodes to be evaluated. We were able to set-up only two server nodes. With two server nodes, the algorithm correctness was tested and we were able to retrieve the trending hashtags from the two server nodes.

The second experiment was performed to test the recommendation mode with similar users. This was done using the dataset obtained from Twitter with its Streaming API. As mentioned before we collected 10000 tweets of trending topics from specifically the ones with some of the politicians tagged in them. We wanted to check the maximum number of hashtags out of the total recommendations that would be relevant for the given content. We distributed around 600 tweets to each of the clients, increasing the number of clients at each step. When there was only one user we were not able to retrieve any related recommendations. At 10 and 15 client count we were able to retrieve 4 relevant hashtags of the recommendations made. Hence, as the users increase we would be able to provide with top-k recommendations with k = 5. Figure 5 shows the results for this experiment.

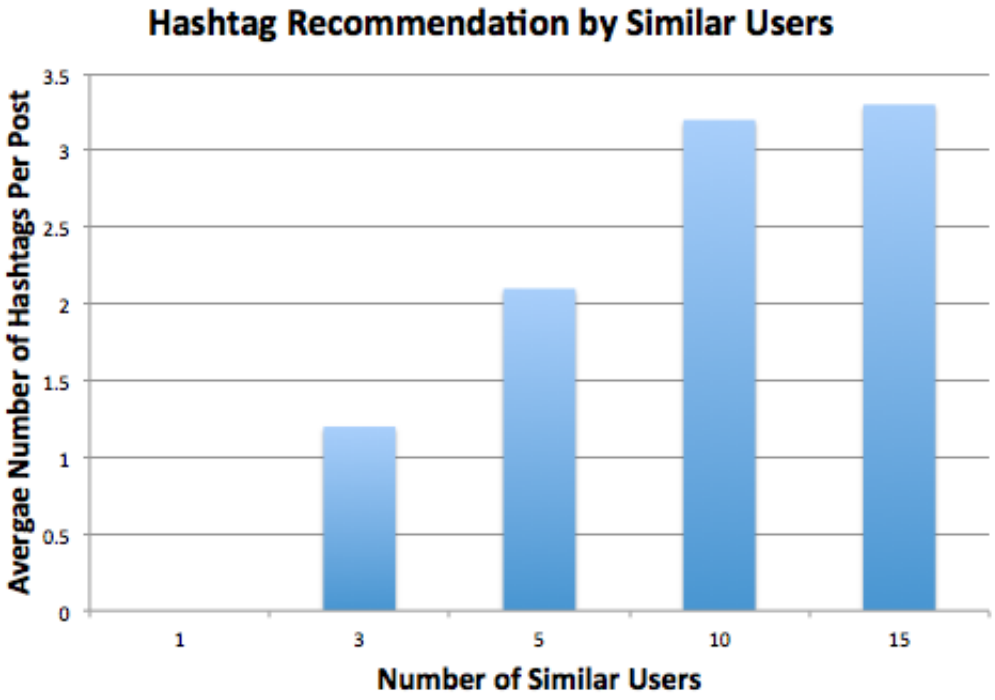


Figure 5: Average hashtags per post Vs Number of similar users

The last experiment was to check hit-rate when we used both user preferences and similar users recommendations at the same time. We used the same data from Twitter Streaming API for this experiment too. We tested for top-k recommendations when k = 5 and k = 10. For hit-1 this choice was good and we were able to acquire around 87% and 92% hit-rate for top-5 and top-10 recommendations respectively. The hit-all rate which checks for a match for all of the hashtags in a tweet, was around 63% when we used top-5 recommendations and it was still below 50% when top-10 recommendations were considered. From the results of this experiment we intuit that may be by combining more than one mode together the proposed approach could provide better results. Table 1 shows the results for the same.

k	Measurement	Values
5	Hit-1	0.87
5	Hit-all	0.63
10	Hit-1	0.92
10	Hit-all	0.49

Table 1: Hit-rate for top-k recommendations with User Preferences and Similar Users

6. CONCLUSION AND FUTURE WORK

In this paper we introduce ourpeer-to-peer social networking architecture and its components. We also propose hashtag recommendation approach proposed for this application using Latent Dirichlet Allocation [3] topic model. It is model, which identifies hidden topics from a set of pre-

processed documents. We specifically concentrate on identifying topic-hashtag distributions out of these hidden topics. These are further used for the recommendations. Our research uses both content-based and collaborative filtering methods for the recommendations, which can be selected by the user on his own choice. Also, we provide the recommendations by considering content from the neighbouring nodes in the network, which would allow us for the fast processing of the recommendations. The experiment results show more than 50% hit-rate for three of the collaborative filtering approaches. The hit-1 rate for top-5 and top-10 recommendations for hashtags considered from similar users and user content is the better than any of the topic model based hashtag recommendation systems. Also, using only similar users method guarantees that the approach is good for top-3 recommendations.

There are some limitations as to the proposed recommendation methodology. We still have to test the performance of the algorithms in peer-to-peer simulated environment with more number of server nodes. Without which we were not able to test the overall popularity method. The next thing would be to consider a top-k recommendation system for all of the methods mentioned. As a future work, we would give the recommendation methods to the user as part of advanced settings and include more than one method for a recommendation. We use relational database for storing both the bootstrap and server peer data. With the users increasing, at some point we need to consider moving to BigData solutions. Also, we need add in encryption mechanisms for securing the client data stored on the server.

ACKNOWLEDGEMENTS

We would like to thank the development team and University of Ottawa for the tremendous support in this research.

REFERENCES

[1] Adomavicius, G., &Tuzhilin, A.(2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions.Knowledge and Data Engineering, IEEE Transactions On, 17(6), 734-749.

[2] Blei, D.M. (2012).Probabilistic topic models.Communications of the ACM, 55(4), 77-84.

[3] Blei, D.M., Ng, A.Y., & Jordan, M.I. (2003).Latent dirichlet allocation.J.Mach.Learn.Res., 3, 993-1022.

[4] Chen, G., Hu, T., Jiang, D., Lu, P., Tan, K., Vo, H. T., et al. (2012). BestPeer++: A peer-to-peer based large-scale data processing platform. Proceedings of the 2012 IEEE 28th International Conference on Data Engineering, pp. 582-593.

[5] Chuang, H., Wang, L., & Pan, C.(2008).A study on the comparison between content-based and preference-based recommendation systems.Semantics, Knowledge and Grid, 2008.SKG '08. Fourth International Conference On, pp. 477-480.

[6] Godin, F., Slavkovikj, V., De Neve, W., Schrauwen, B., & Van de Walle, R. (2013). Using topic models for twitter hashtag recommendation. Proceedings of the 22nd International Conference on World Wide Web Companion, pp. 593-596.

[7] Griffiths, T.L., &Steyvers, M.(2004). Finding scientific topics. Proceedings of the National Academy of Sciences, 101(suppl 1), 5228-5235.

[8] Jagadish, H.V., Ooi, B.C., & Vu, Q.H. (2005). BATON: A balanced tree structure for peer-to-peer networks. Proceedings of the 31st International Conference on very Large Data Bases, Trondheim, Norway. pp. 661-672.

[9] Jeon, M., Jun, S., & Hwang, E. (2014).Hashtag recommendation based on user tweet and hashtag classification on twitter.Web-age information management (pp. 325-336) Springer.

[10] Krestel, R., Fankhauser, P., &Nejdl, W. (2009).Latent dirichlet allocation for tag recommendation.Proceedings of the Third ACM Conference on Recommender Systems, New York, New York, USA.pp. 61-68.
[11] Leskovec, J., Rajaraman, A., & Ullman, J. D. (2014).Mining of massive datasets Cambridge University Press.
[12] Lops, P., De Gemmis, M., &Semeraro, G. (2011). Content-based recommender systems: State of the art and trends. Recommender systems handbook (pp. 73-105) Springer.
[13] McFedries, P. (2013). Hashtag, you're it [technically speaking].Spectrum, IEEE, 50(4), 24-24.
[14] Peng, X., Cao, Y., &Niu, Z.2008).Mining web access log for the personalization recommendation.MultiMedia and Information Technology, 2008.MMIT '08. International Conference On, pp. 172-175.
[15] Potts, L., Seitzinger, J., Jones, D., & Harrison, A. (2011). Tweeting disaster: Hashtag constructions and collisions. Proceedings of the 29th ACM International Conference on Design of Communication, Pisa, Italy.pp. 235-240.
[16] She, J., & Chen, L. (2014). TOMOHA: TOpic model-based HAshtag recommendation on twitter. Proceedings of the Companion Publication of the 23rd International Conference on World Wide Web Companion, Seoul, Korea.pp. 371-372.
[17] Steyvers, M., & Griffiths, T.(2007).Probabilistic topic models.Handbook of Latent Semantic Analysis, 427(7), 424-440.
[18] Wang, S. (2014).uopStore: An ecommerce platform with a peer-to-peer infrastructure. School of Electrical Engineering and Computer Science, University of Ottawa).
[19] Yu, J., &Shen, Y. (2014).Evolutionary personalized hashtag recommendation.Web-age information management (pp. 34-37) Springer.
[20] Zangerle E., G. W. Recommending hashtags in twitter.Workshop on Semantic Adaptive Social Web 2011, in Connection with the 19th International Conference on User Modeling, Adaptation and Personalization (2011).

INTENTIONAL BLANK

ENHANCEMENT AND SEGMENTATION OF HISTORICAL RECORDS

Soumya A[1] and G Hemantha Kumar[2]

[1]Dept. of Computer Science & Engg,
R V College of Engineering, Bangalore, India
soumyaa@rvce.edu.in
[2]Dept. of Studies in Computer Science,
University of Mysore, Mysore, India
ghk.2007@yahoo.com

ABSTRACT

Document Analysis and Recognition (DAR) aims to extract automatically the information in the document and also addresses to human comprehension. The automatic processing of degraded historical documents are applications of document image analysis field which is confronted with many difficulties due to the storage condition and the complexity of the script. The main interest of enhancement of historical documents is to remove undesirable statistics that appear in the background and highlight the foreground, so as to enable automatic recognition of documents with high accuracy. This paper addresses pre-processing and segmentation of ancient scripts, as an initial step to automate the task of an epigraphist in reading and deciphering inscriptions. Pre-processing involves, enhancement of degraded ancient document images which is achieved through four different Spatial filtering methods for smoothing or sharpening namely Median, Gaussian blur, Mean and Bilateral filter, with different mask sizes. This is followed by binarization of the enhanced image to highlight the foreground information, using Otsu thresholding algorithm. In the second phase Segmentation is carried out using Drop Fall and WaterReservoir approaches, to obtain sampled characters, which can be used in later stages of OCR. The system showed good results when tested on the nearly 150 samples of varying degraded epigraphic images and works well giving better enhanced output for, 4x4 mask size for Median filter, 2x2 mask size for Gaussian blur, 4x4 mask size for Mean and Bilateral filter. The system can effectively sample characters from enhanced images, giving a segmentation rate of 85%-90% for Drop Fall and 85%-90% for Water Reservoir techniques respectively.

KEYWORDS

Document Analysis, Preprocessing, Filters, Segmentation, Drop Fall Technique, Water Reservoir Technique

1. INTRODUCTION

A generic Optical Character Recognition (OCR) system comprises of different stages like preprocessing, segmentation, feature extraction and classification. Preprocessing is one of the most interesting and challenging topics in DAR. Preprocessing of document involves converting scanned images or photographed images of machine printed or handwritten text which may

David C. Wyld et al. (Eds) : ACITY, DPPR, VLSI, WiMNET, AIAA, CNDC - 2015
pp. 95–113, 2015. © CS & IT-CSCP 2015 DOI : 10.5121/csit.2015.51309

include numbers, letters and symbols into system processable format. Segmentation is an important assignment of any OCR system and it separates the image text documents into lines,words and characters. Hence the accuracy of OCR system primarily depends on the segmentation algorithm been used.

Segmentation of handwritten text of Indian languages is challenging when compared with Latin based languages because of its structural complication and presence of compound characters. This complexity increases further if were to recognize text of ancient Indian or non-Indian epigraphical documents. The epigraphical records engraved on stones, rocks, pillars or on some other writing material are non-linear in their shapes and non-uniform in their sizes.

Raw image of an epigraph contains unwanted symbols or marks, noise embedded and text engraved with much skew.The spacing between characters and also between the lines and the skew could complicate the process of translating the scripts. Some touching lines as well as characters complicates the process of segmentation which is input for the recognition process in the later stages. Hence the input document image of epigraphs is to be preprocessed for removal of noise, skew detection and correction, followed by segmentation of characters [1].

Inspite of several positive works on OCR across the world, development of OCR tools in Indian languages is still a challenging task. Character segmentation plays an important role in character recognition since incorrectly segmented characters are susceptible to be recognized wrongly. Hence the proposed work focuses on preprocessing and segmentation of ancient handwritten documents. This is an initial step towards developing OCR for ancient scripts, which can be used by archaeologists and historians for digitization and further exploration of ancient records.

This paper is organized as follows: Section 2 elaborates the related works in the field. The system architecture is highlighted Section 3. The theory and related mathematical background of the approaches in current system is discussed in Section 4. Methodology is given in Section 5. Experimental results and performance analysis is covered in Section 6 and Section 7 provides conclusion.

2. RELATED WORK

Researchers have worked on many approaches for preprocessing and segmentation of various languages. In this section, some of the works are discussed.

The linear Unsharp Masking (USM) technique [2] is adopted to increase the pictorial presence of an image by highlighting its regularity contents to improve the edge and detailed information in it. Nevertheless this method is easy and gives good result for many applications. It has two limitations, one it is tremendously sensitive to noise and other one is that it increases high contrast areas much more than area that do not show high image dynamics. Therefore output image suffers from unkind overshoot objects. Adaptive Unsharp Masking hires an adaptive filter that controls the contribution of sharpening in the manner that contrast enhancement occurs in high dense areas and less or no image sharpening occurs in soft areas. This algorithm is better compared with several other methods available in linear unsharp masking filter technologies.

Binarization is the initial step for processing, with the fact of degradation of the source document, whichever global or local thresholding approaches are chosen. The Otsu thresholding algorithm

[3] using the histogram shape analysis, which is most widespread global binarization algorithm. The thresholding of Otsu yields a promising performance when the histogram has twin modal distribution. The global threshold is designated automatically by a discriminant standard.

Another method utilize image contrast defined as local image minimum and maximum when compared with the image gradient process, the image contrast derived by the local maximum and minimum process has a good property and it is more tolerant to the uneven illumination, document degradation such as smudge [4]. This method is superior when handling document images with difficult background variation. Finally, the ancient document image is binarized based on the local thresholds that are derived from the detected high contrast image pixels when the same is compared with previous method based on image contrast, the method uses the image contrast to recognize the text stroke boundary and it can be used to produce high accurate binarization results.

A common technique for scrubbing the degraded documents is modified iterative global threshold algorithm [5]. A best approach in the separation of object information from foreground is to compute a global threshold of intensity value based on which two clusters can be diverted. It is an iteration approach which can handle many degraded conditions. In each iteration the intermediate tones are shifted towards background there by providing efficient difference between foreground and background. It is mostly useful for the documents having non-uniform distribution of noises.

In order to make foreground inscriptions clearly visible from background, Histogram normalization is used. The image obtained still may suffer from uneven background intensity variation which in turn reduces the clarity of the foreground. Further processing of image is done to get an image with better foreground information. As the intensity of the foreground pixels differs from the intensity of background, this key factor is used to identify the foreground characters. The main criteria is to find a threshold value which causes the image components to lie in one of two levels L0, which is below the threshold value and L1, which is above the threshold value. The pixels which are above the threshold value represent the nodes of a graph. The nodes form the basis for representing the foreground. The nodes that are neighbors in the sampling grid are joined by an edge [6].

This technique is used to reduce the number of pixels in the image by a factor of 4 or 8, which in turn will decrease the number of pixels that has to be processed. The image is represented as a graph by associating each pixel to a vertex of a graph and connecting the pixels that are neighbors in the sampling grid by an edge. The gray value of the pixel is considered as an attribute of the vertex. Since the image is of finite size so also the graph. Pixels represent the finite regions and vertices represent the faces. The dual of this graph represents borders of the faces which are inter-pixel edges and vertices. Dual graph pyramids are constructed by adopting bottom-up approach. Each level of pyramid represents an adjacency graph where vertices correspond to regions and edges represent the relation between the regions [7].

The procedure of segmentation has huge importance in the handwritten script identification. Thus an abundant study of research outcome in related segmentation field was surveyed. The algorithm based on connected components [8], segmenting the document image into non-overlapping equi-width vertical zones.

A method is to decrease the noise level, which is existing in the distorted image is proposed in [9]. The distorted image will be first binarized using the threshold, which is determined by the Otsu's method. Next for each pixel p(x, y) estimate the horizontal and vertical run length count. If horizontal and vertical run length count is less than a specified threshold then it is assumed to be noise and will be eliminated.

Gaussian kernel is a linear operation; convolution is used to find the common area between the profile and the Gaussian kernel. The degree of shift in the Gaussian kernel during the convolution process linearly varies with horizontal profile information. So this can be used to represent randomness in the profile and provides a zero crossing smooth curve, when it is convolved with the profile, represented by 'C'. The peaks which are above zero are treated as the gaps between the lines. Based on this information, the line segmentation is performed [10].

Another method for segmentation is nearest neighbor algorithm [11] which is iterative in nature scans the character from the top left portion of the image. When it reaches a first black pixel, then the first symbol is identified through the connected component. If it is found to be the first character of the script, then it will be placed as the first character of the new line used for placing the character segmentation. The centroid of the character is calculated and stored in an array separation the x and y coordinators. The document is again scanned from left top to locate the next black pixel and hence the next character in the document. The centroid of the character also computed. The distance between the centroid is computed using the distance formula. If the distance is less than or equal to threshold value then the character is assume to lie on same line. Otherwise the character is consider being the part of the next line and transferred to the next line in the result part. This process is continued until all the characters are scanned and whole image has been traversed. At the end of the iterative algorithm the separated lines are obtained from the source. The individual character can also be obtained in this process itself.

Text line segmentation is necessary to detect all text regions in the document image. The algorithm based on multiple histogram projections using morphological operators to extract features of the image. Horizontal projection is performed on the text image, and then line segments are identified by the peaks in the horizontal projection. Threshold applied to divide the text image into segments. False lines are eliminated using other threshold. Vertical histogram projections are used for the line segments and decomposed into words using threshold and further decomposed to characters. This kind of approach provides best performance based on the experimental results such as Detection rate DR (98%) and Recognition Accuracy RA (98%) [12,13].

Contour tracing is a technique applied to digital images for extracting the boundary of any object. This kind of system applies one of the recent contour tracing algorithms to separate character by using the Theo Pavlidis's algorithm [15]. It works with 4-connected patterns. The width of the segmented components from this process is checked. If it is more than the criteria value of the average width of the components, it will be processed in the next stage. This means there are some touching characters that are not separated [14].

Tracing of Background Skeleton approach, is applied to separate some touching components to segment touching characters, background skeleton is processed by using the Zhang-Suen thinning algorithm. Then, contour tracing algorithm is applied to abstract the skeleton of the background. Subsequently, the characters in each line will be sorted by checking the column position in order

to determine the sequence of the characters. This is applied to practical data from ancient documents [15, 16].

Enhanced stroke filter have shifted the attention to the skeletons of potential strokes. In this manner, the task of distinguishing strokes from a complicated background is converted to the task of comparing the difference between skeletons of potential strokes and those of disturbing patches, both of which can be extracted from the resulting images of previous Stroke filters. Skeleton constraints such as length and width constraints can be introduced into the method to enhance stroke information [17].

For the segmentation of unconstrained handwritten connected numerals, Water Reservoir technique is used. A reservoir location and size, touching position (top, middle or bottom) is decided. Analyzing the reservoir boundary, touching position and topological structures of the touching pattern, the best cutting point and then the cutting path for segmentation is generated [18].

Segmented linked characters are critical preprocessing steps in character recognition applications. Old drop fall algorithm has proved to be an efficient segmenting method due to its simplicity and effectiveness. However it is subject to small convexes on the contour of characters. Xiujuan Wang, Kangfeng Zheng, and Jun Guo presented Innertial Drop fall algorithm and big drop fall algorithm to avoid this defect [19, 20, 21].

3. SYSTEM ARCHITECTURE

The system "Enhancement and Segmentation of Historical Records" designed, mainly consists of the subcomponents - Preprocessing and Segmentation as shown in Figure 1. The input to the system is ancient epigraphic documents of varying amount of degradation.

- **Image Enhancement :** This sub system enhances the quality of the ancient document images by reducing noise. This is carried out by providing four different filtering options of various filter sizes.

- **Binarization:** This sub-system converts RGB images to binary images using thresholding technique known as Otsu algorithm.

- **Segmentation:** This sub-system samples out characters from the ancient documents and is achieved through Drop Fall and Water reservoir techniques.

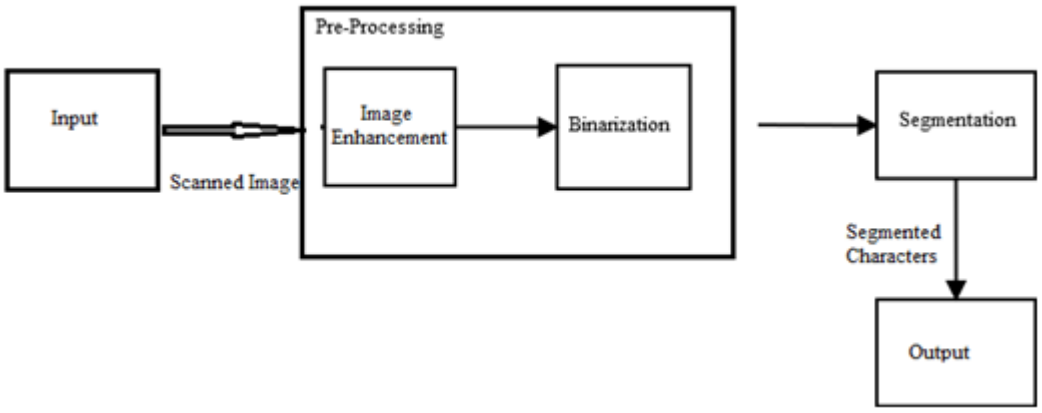


Figure 1. Enhancement and Segmentation System

4. REVIEW ON THE APPROACHES USED IN PROPOSED SYSTEM

The methods used in current system are described in this section:

Preprocessing stage of degraded ancient document images includes: enhancement and reduction in noise, which is achieved through different filtering methods for smoothing or sharpening namely Bilateral, Mean, Median, and Gaussian Blur Filters. These filters are provided with different mask sizes and parameter values. This is followed by binarization of the enhanced image to highlight the foreground information, using Otsu thresholding algorithm. Character segmentation is performed based on Drop Fall and Water Reservoir concept.

4.1 Mean Filter

The mean filter is a sliding-window longitudinal filter that exchanges the center value in the window with the average (mean) of all the pixel values in that window [22]. Let Sxy represent the set of coordinates in a rectangular sub image window of size m x n, centered at point (x, y). The arithmetic mean filtering process computes the average value of the corrupted image g(x, y) in that area defined by Sxy. The restored image value at any point (x, y) is merely the arithmetic mean calculated using the pixel in that region, indicated in Equation 1.

$$\tilde{f}(x,y) = \frac{1}{m^n} \sum_{(s,t)\in S_{xy}} g(s,t) \quad (1)$$

This operation can be applied using a convolution mask in which all coefficients have value 1/mn. Mean Filters smoothes local variations in an image and as a result of blurring, noise is reduced.

4.2 Median Filter

In image processing, neighborhood averaging is the best method to perform the noise reduction, whereas the method can overturn isolated out of range noise, however the adverse effect is that it also distorts sudden changes such as sharp edges. The median filter can suppress the noise without damaging the sharp edges. In median filtering, all the pixel values are first sorted into numerical order and then replaced with the middle pixel value [22].

Let y be a pixel location and w a neighborhood centered on location (m, n) in the image, therefore median filter is given by Equation 2,

$$y[m, n]=\text{median}\{x[i, j],\ (i, j) \text{ belongs to } w\} \quad (2)$$

Subsequently the pixel y [m, n] represents the location of the pixel y, m and n represents the x and y co-ordinates of pixel y. w represents the neighborhood pixels surrounding the pixel position at (m, n), (i, j) belongs to the same neighborhood centered on (m, n). Hence the median method will take the median of all the pixels within the range of (i, j) represented by x [i, j].

4.3 Gaussian blur Filter

A Gaussian blur or Gaussian smoothing involves blurring an image by Gaussian function and used to decrease image noise and image details. Gaussian smoothing is used as a preprocessing stage in computer vision algorithms in order to enhance image structures at different scales. The 2- dimension Gaussian function is given by Equation 3.

$$G(x, y) < -\frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (3)$$

where x is the distance from the origin (Horizontal axis), y is the distance from the origin (vertical axis), and σ is the standard deviation of the Gaussian distribution. The standard deviation σ of the Gaussian determines the amount of smoothing [22].

4.4 Bilateral Filter

Bilateral filter [23] is a non linear filter in spatial domain, which does averaging without smoothing the edges. The bilateral filter inputs a weighted sum of the pixels in a local neighborhood; the weights depend on both the spatial distance and the intensity distance. Essentially the bilateral filter has weights as a product of two Gaussian filter weights, one of which corresponds to average intensity in a spatial domain, and second weight corresponds to the intensity difference. Hence no smoothing occurs, when one of the weights is close to 0, which means the product becomes insignificant around the region where intensity changes swiftly, which represents usually the sharp edges. As a result, the bilateral filter preserves sharp edges [28]. Pixel location x, bilateral filter output is given in Equation 4

$$\grave{I} = \frac{1}{C}\sum_{y\in N(x)} e^{\frac{-||y-x||^2}{2\sigma_d^2}} - e^{\frac{-|I(y)-I(x)|^2}{2\sigma_r^2}} \quad I(y) \quad (4)$$

There parameters controlling the fall-off of weights in spatial and intensity domains, respectively. And are inputs and output images respectively are spatial neighborhood of pixel I(x), and C can be given as

$$C = \sum_{y\in N(x)} e^{\frac{-||y-x||^2}{2\sigma_d^2}} - e^{\frac{-|I(y)-I(x)|^2}{2\sigma_r^2}} \quad (5)$$

4.5 Otsu's Method of Binarization

Binarization is the method of converting a grey scale image to a binary image by using threshold selection procedures to categorize the pixels of an image into either one of the two classes. Binarization of the image using Otsu method [23] is used to automatically accomplish histogram shape-based image thresholding or the decrease of a gray level image to a binary image. This algorithm adopts that the image as thresholded contains two classes of pixels, then calculates the

optimum threshold separating those two classes so that their combined spread (intra-class variance) is minimal.

In an Otsu's method weighted sum of variances of the two classes is given by:

$$\sigma_{\omega}^2(t) = \omega_1(t)\,\sigma_1^2(t) + \omega_2(t)\,\sigma_2^2(t) \tag{6}$$

Weights ω_i are the probabilities of the two classes separated by a threshold t and σ_i^2 variances of these classes. The class probability is

$$\sigma_b^2(t) = \sigma^2 - \sigma_{\omega}^2(t) = \omega_1(t) = \omega_1(t)\omega_2(t)\,[\mu_1(t) - \mu_2(t)]^2 \tag{7}$$

The class probability $\omega_1(t)$ is calculated from the histogram as t:

$$\omega_1(t) = \sum_0^t p(i) \tag{8}$$

While the class mean $\mu_1(t)$ is:

$$\mu_1(t) = [\sum_0^t p(i)x(i)] \;/\; \omega_1 \tag{9}$$

where x(i) is $\omega_2(t)$ the value at the center of the ith histogram bin. Also can calculate μ_2 on the right-hand side of the histogram for bins greater than t.

4.6 Drop Fall Algorithm for Segmentation

Drop fall algorithm [24] with respect to the principle that an equally ideal cut between two touched characters can be created, if one has to role a hypothetical marble off the top of the first character and create the cut where the marble falls. The important things to be addressed for this implementation are where to drop the marble from because it is important if the algorithm starts at the wrong place. The marble can simply roll down the left side of the first digit or the right side of the second digit and, hence, it would be completely unsuccessful. The best approach to start drop falling process is possible to the point at which two characters are touched. In this process the pixels are scanned row by row until a black boundry pixel with adjacent black boundry pixel to the right of it is identified, where as the two pixels are separated by white space. This pixel is used as a point to start the drop fall as shown in Figure 2.

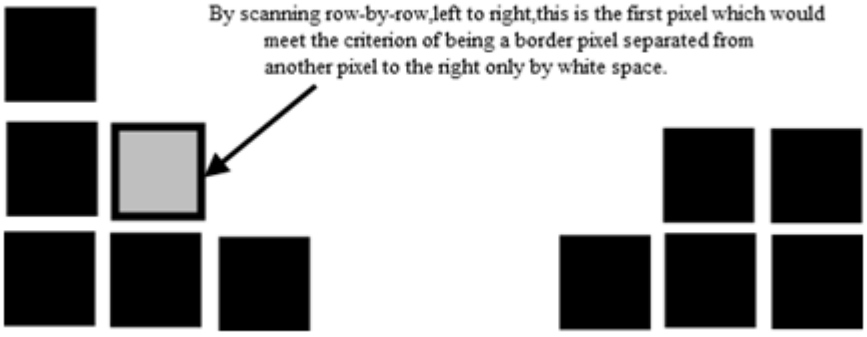


Figure 2. Identification of Initial Pixel Positions

The direction that the algorithm will move is according to the current pixel position and its surroundings as shown in Figure 3.

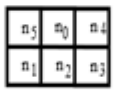


Figure. The neighborhood pixels of n 0

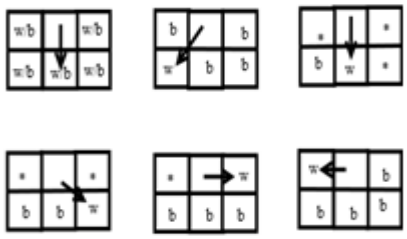


Figure. The principle of drop-fall algorithm

Figure 3. The Principle of Drop Fall algorithm

4.7 Water Reservoir Algorithm

The larger space generated by touching characters is analyzed with the help of water reservoir concept. The working principle of water reservoir method is illustrated in Figure 4. When water is poured from top (bottom) of a component, the regions of the component where water will be stored are considered as top (bottom) reservoir. Top (bottom) reservoir is the reservoir obtained when water is poured from top (bottom). The white spaces are found in the regions in the bounding box of the components where water can be stored. These regions are called water reservoirs. The reservoirs obtained in this procedure are not considered for further processing. Those reservoirs whose heights are greater than a threshold value T1 are considered for further processing.

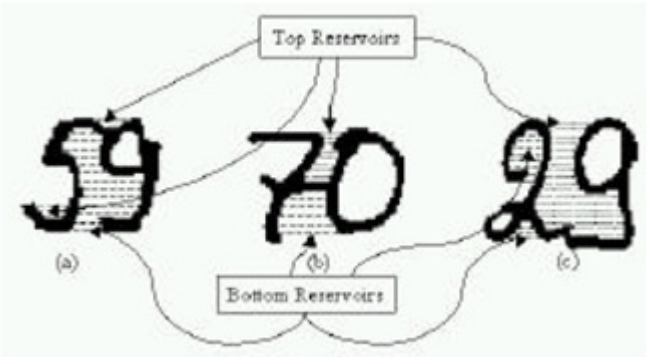


Figure 4. Reservoirs formed from water flow from top and bottom is shown for (a) top (b) middle and (c) bottom touching numerals. The top Reservoirs are marked by Dots and bottom reservoirs are marked by small line segments

When two characters touch each other, they create a space (reservoir) between the characters. This space is very important for segmentation because,

a) As cutting points are concentrated around the base of the reservoir, and hence, decreases the search area.

b) The cutting points lie on base of the reservoir.

c) The space attributes (center of gravity and height) aid to go near the best touching position.

If water is poured from top (bottom) of large space created by touching (Water reservoir) Base of the reservoir connected numeral then water will be stored in this large space. This water stored area is named "Water Reservoir" [25]. Figure 5 illustrates the same.

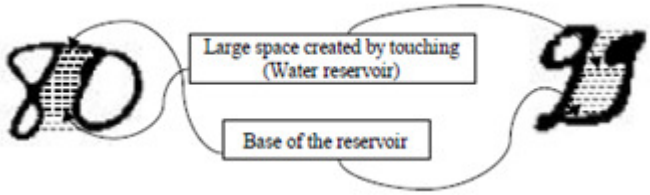


Figure 5. Examples of touching numeral and Space created by the touching

Figure 6 depicts the detection of touching position recognition. The largest reservoir of the component whose center of gravity lies in vm region is found. This reservoir is known as the best reservoir for touching. The base-line (lowermost row of the reservoir) of the best reservoir is then identified. The best reservoir and its base-line are shown and to find this touching position in the components, morphological thinning operation is applied to touching components for further processing. For feature points extraction the touching position is renowned. The leftmost and rightmost points of the base-line of considered reservoirs are the feature points. These points are initial feature points. With this initial feature points the best feature point (which gets maximum confidence value) is chosen for segmentation. To calculate confidence value (CV) following features are considered. Euclidean distance of feature points from the center of gravity of the touching component.

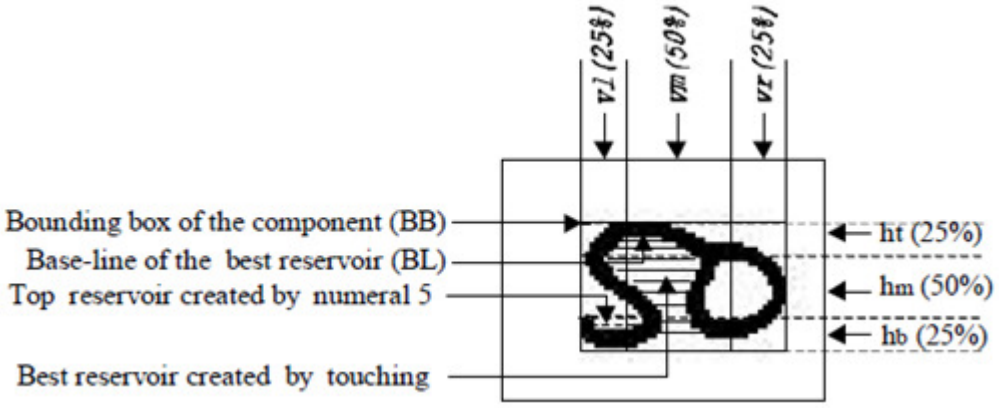


Figure 6. Feature Detection Approach.

5. PROPOSED SYSTEM AND METHODOLOGY

5.1 DFD of the Preprocessing and Segmentation System

The Data Flow Diagram(DFD) of the current system is shown in Figure 7. This work is carried out in two phases. In the first phase - Preprocessing, the degraded ancient document image is taken as input and it is converted to grayscale. Then the smoothing or sharpening filters, namely Median filter, Gaussian filter, Mean filter, Bilateral filter of different mask sizes are applied to reduce the amount of noise and thus enhances the image. Next, the enhanced image is binarized using Ostu algorithm to differentiate background and foreground of ancient document.

In the second phase Segmentation is carried out using Drop Fall algorithm and Water Reservoir algorithm, to obtain sampled characters, which can be used later stages of OCR.

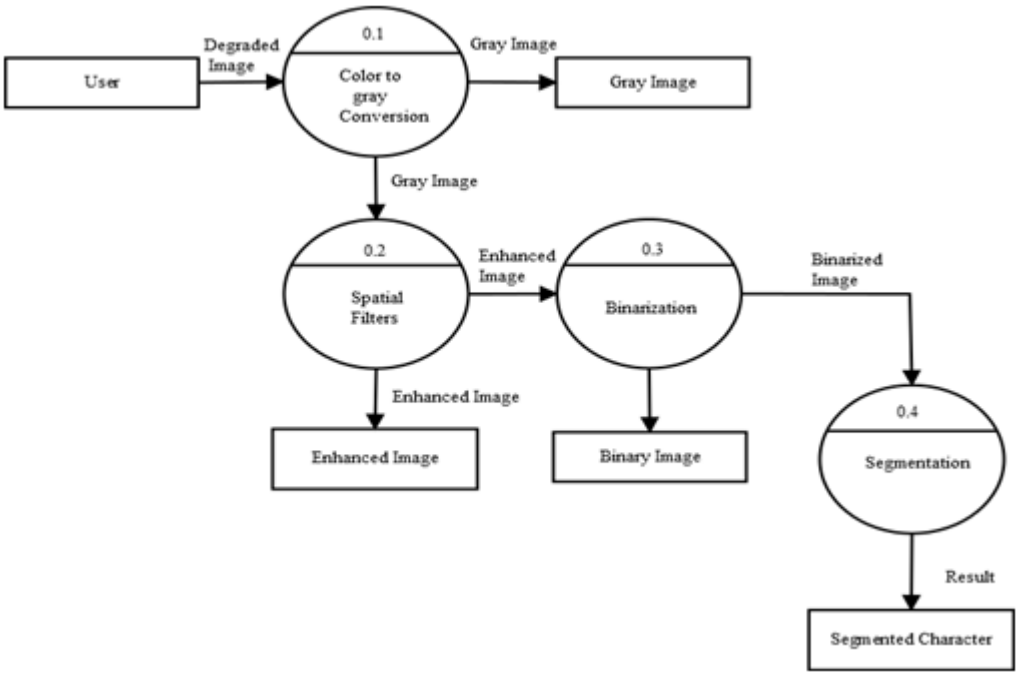


Figure 7. DFD of the Preprocessing and Segmentation Phases

5.2 Methodology

The functionality of the phases – Preprocessing and Segmentation in detail is covered in this section. The ancient input image is first converted to grayscale image. The image is enhanced and noise is reduced by applying four different filters with mask size of 2x2 and 4x4. Next, The enhanced image is converted to binary image. Lastly the characters in the document are sampled out using Drop Fall and Water Reservoir approaches.

5.2.1 Preprocessing

- **Image Enhancement**

  - **Input:** Degraded Gray Scale image
  - **Functionality:** Enhances epigraphic image of medium-level degradation using spatial filters namely Median, Gaussian blur, Mean, Bilateral filter on the input image.
  - **Output:** The enhanced image with reduced noise.

  - **Algorithm for Enhancement using Median filter**

    [Step 1]: Read the gray image.

    [Step 2]: Compute y [m, n] = median{x [i, j], (i, j) belongs to w}

    y be a pixel position, w represent a neighborhood centered around location (m, n) in the image.

    [**Step 3]:** Apply the Median filter designed over the entire input image to obtain enhanced image.

  - **Algorithm for Enhancement using Gaussian blur filter**

    [Step 1]: Read the gray image.

    [Step 2]: Compute$e$ $\mathrm{G}(\mathrm{x},\mathrm{y}) < -\frac{1}{2\pi\sigma^2}\mathrm{e}^{-}\frac{\mathrm{x}^2+\mathrm{y}^2}{2\sigma^2}$

    x is the distance from the origin (Horizontal axis), y is the distance from the origin (vertical axis), σ is the standard deviation of the Gaussian distribution.

    [Step 3]: Apply the Gaussian filter designed over the entire input image to obtain enhanced image.

  - **Algorithm for Enhancement using Mean filter**

    [Step 1]: Read the gray image.

    [Step 2]: Compute $\hat{f}(x,y) < -\frac{1}{m^n}\sum_{(s,t)\epsilon S_{xy}} g(s,t)$

    Sxy represent the set of coordinates in a rectangular subimage window of size m X n, centered at point (x, y).

    [Step 3]: Apply the Mean Filter designed over the entire input image to obtain enhanced image.

  - **Algorithm for Enhancement using Bilateral filter**

    [Step 1]: Read gray image.

    [Step 2]: Compute $C = \sum_{y\in N(x)} e^{\frac{-||y-x||^2}{2\sigma_d^2}} - e^{\frac{-|I(y)-I(x)|^2}{2\sigma_r^2}}$

    [Step 3]: Compute $\grave{I} = \frac{1}{C}\sum_{y\in N(x)} e^{\frac{-||y-x||^2}{2\sigma_d^2}} - e^{\frac{-|I(y)-I(x)|^2}{2\sigma_r^2}} \quad I(y)$

    [Step 4]: Apply the Bilateral Filter designed over the entire input image to obtain enhanced image.

- **Binarization**
  - **Input:** Enhanced Image
  - **Functionality:** The enhanced images are converted to binary image consisting of ones and zeroes.

- **Output:** Binarized image
- **Algorithm for Binarization**
  [Step 1]: Compute histogram and probabilities of each intensity level
  [Step 2]: Initialize initial $\omega_i(0)$ and $\mu_i(0)$
  [Step 3]: Compute through all possible thresholds t=1. Maximum intensity
  Revise ω_i and μ_i; Compute σ_b^2(t)
  [Step 4]: Desired threshold corresponds to the maximum σ_b^2(t)
  [Step 5]: Compute two maxima (and two corresponding thresholds). σ_{b1}^2(t)
  [Step 6]: Compute greater max and σ_{b2}^2(t) is the greater or equal maximum
  [Step 7]: Compute required threshold ← threshold1+threshold2/2

5.2.2 Segmentation

The segmentation of the document image is carried out at the character level using Drop Fall and Water Reservoir Approaches.

- **Input:** Binary epigraph image
- **Functionality:** The binarized image is segmented to characters
- **Output:** Segmented characters of the input epigraph.

➢ **Drop fall algorithm**

Drop falling algorithm forms segmentation path by rolling in between two touching characters and displays the segmented characters.

[Step 1]: Input the binary image
[Step 2]: Find the Height and Width of the touched characters
[Step 3]: Apply Breadth First Search (BFS) algorithm to find the touched characters
[Step 4]: If found start the Drop fall, the drop falling algorithm it will always move downwards, crossways down-wards, to the right, or two the left.
[Step 5]: Make the slice where marble parks. Thus Segmentation path for connected components is found

➢ **Water Reservoir Algorithm**

[Step 1]: Find the size of the characters to find touched characters.
[Step 2]: The positions and sizes of the reservoirs are analyzed and a reservoir is detected where touching is made, the initial feature points for segmentation are noted.
[Step 3]: The best feature points are noted from the initial feature points.
[Step 4]: Based on touching position, close loop positions and morphological structure of touching region the cutting path is produced.

6. EXPERIMENTAL RESULTS, ANALYSIS AND DISCUSSION

6.1 Experimental Results

The system developed is tested on nearly 150 ancient epigraphic images and the results are found to be satisfactory. The sample experimental results are depicted in following figures. Figure 8 shows the input ancient historical record, which is analysed for different spatial filtering techniques.

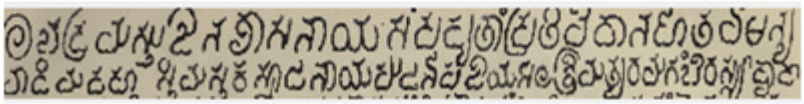


Figure 8. Input Image selected for Pre-processing

Figure 9 shows the results of color to gray scale conversion, when carried out on the image shown in Fig 8.

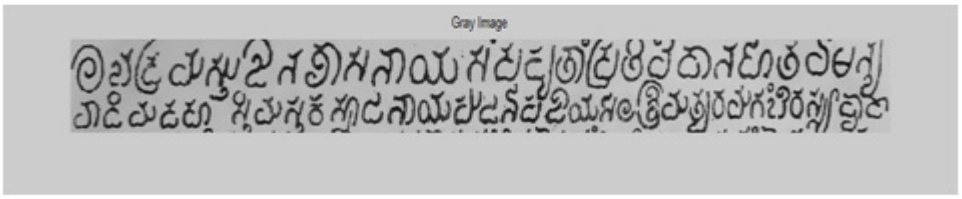


Figure 9. The result of Gray Scale Conversion

Figure 10(a) and 10(b) shows the results after Median filtering, for the mask size of 2x2 and 4x4 respectively. The median filter is an effective method that can suppress isolated noise without blurring sharp edges.

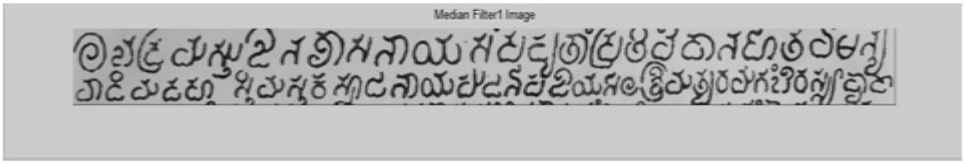


Figure 10(a). The result of Median filtering for Mask size 2x2

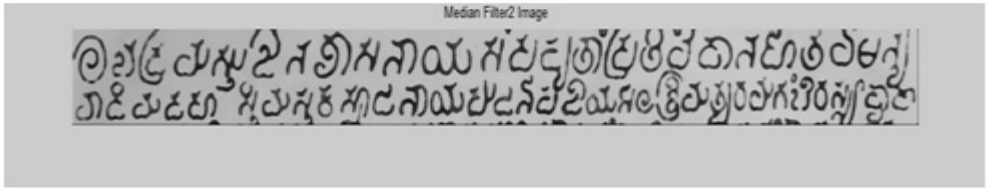


Figure 10(b). The result of Median Filter for Mask size 4x4

Figure 11(a) and 11(b) shows the results of Gaussian blur filtering for the mask size of 2x2 and 4x4. The Gaussian blur method is used to blur the sharpen image so that a less edge highlighted image is produced.

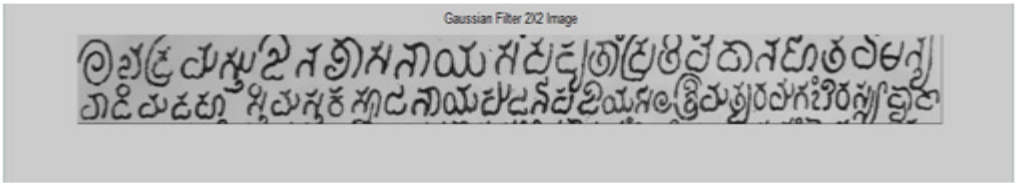


Figure 11(a). The result of Gaussian Blur Filtering for Mask size 2x2

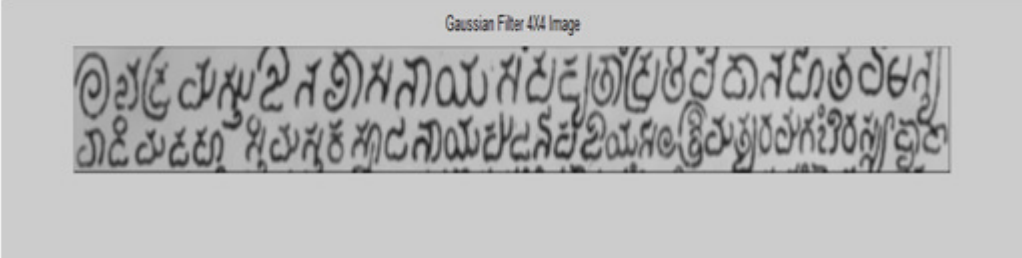


Figure 11(b). The result of Gaussian blur Filtering for Mask size 4x4

Figure 12 shows the results of Mean filtering for the mask size of 4x4. The Mean filter is a simple filter that replaces the center value in the window with the mean of all the pixel values in that window.

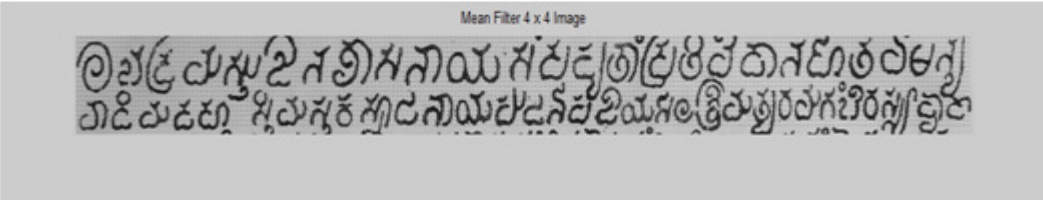


Figure 12. The result of Mean Filtering for Mask size 4x4

Figure 13 shows the results of bilateral filtering.

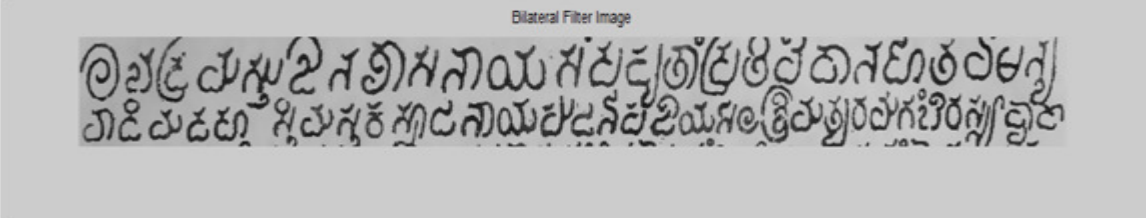


Figure 13. The result of Bilateral Filtering

Figure 14 shows the result of binarization, in which the enhanced image is converted to binary image.

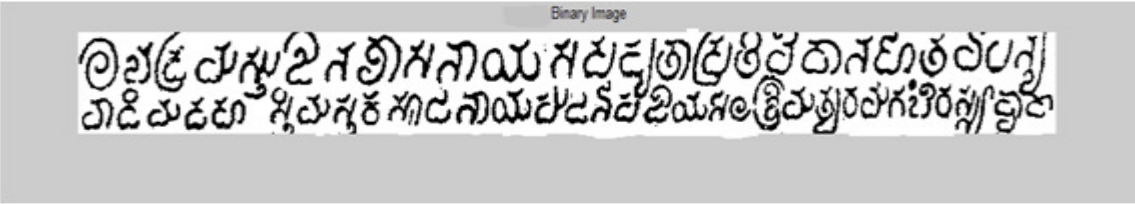


Figure 14. The results of Binarization

Figure 15 and Figure 16 represents the result of Segmentation of Characters using Drop Fall algorithm and Water Reservoir algorithm respectively.

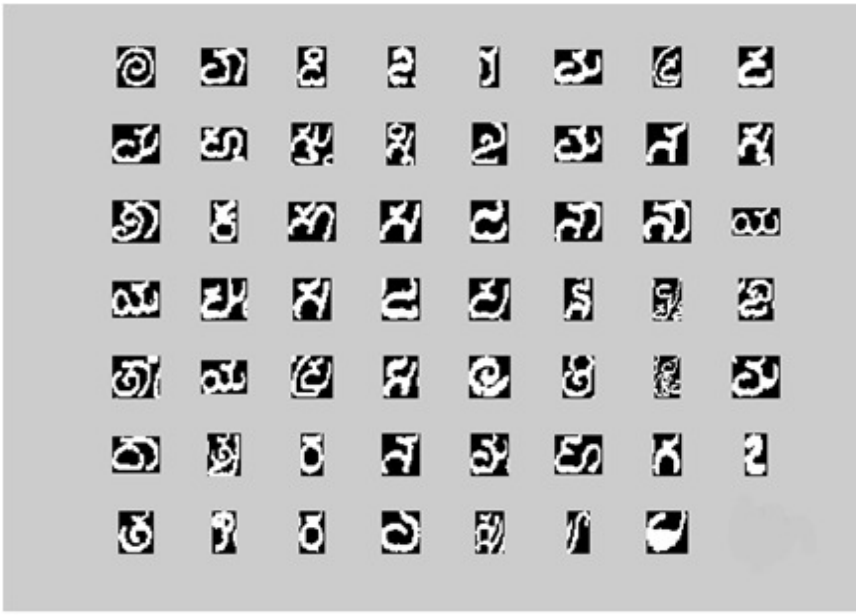


Figure 15. The result of Segmented Characters using Drop Fall algorithm

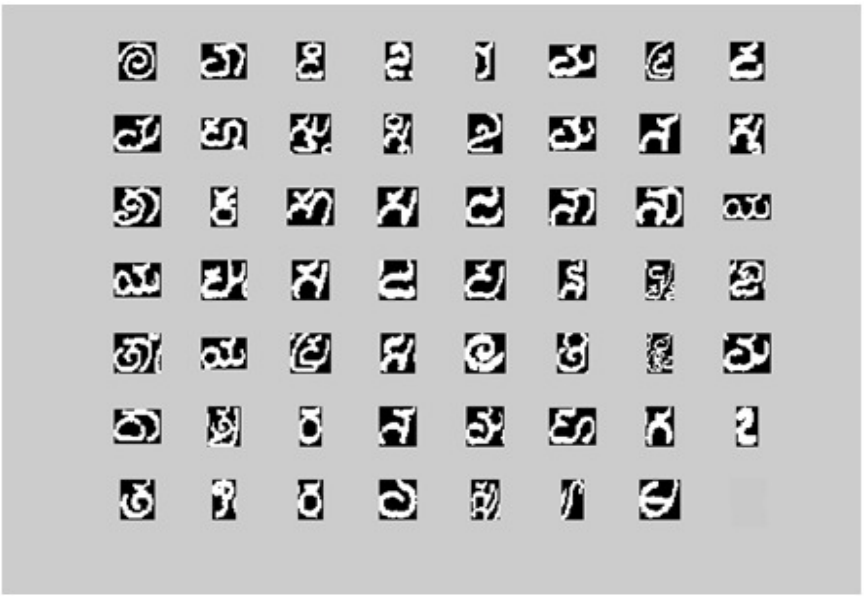


Figure 16. The result of Segmented Characters using Water Reservoir algorithm

6.2 Performance Analysis

The dataset includes 150 samples of medium degraded images for preprocessing and segmentation. The performance of the 2 phases is discussed below:

6.2.1 Filtering Techniques

This system is tested on 150 samples of medium degraded images, using four spatial filtering techniques of varying mask sizes. The enhancement was found to be appreciable for the mask size 4x4 for Median filter when the mask size is high then the output image will appear clear with sharp edges. The mask size of 2x2 for Gaussian blurs results in blurred image. The mask size of 4x4 for Mean filter typically smoothens local variations in an image and noise is reduced as a result of blurring. Bilateral filter sharpens the edges. The smoothing Gaussian filter will result in good accuracy if the edge of the input image is very thick, where as in case of sharpening Bilateral filter gives better output for the medium degraded images.

6.2.2 Segmentation

The system showed good results when tested on 150 varying degraded images, giving segmentation rate of 85%-90% for Drop Fall algorithm, 85%-90% for Water Reservoir algorithm.

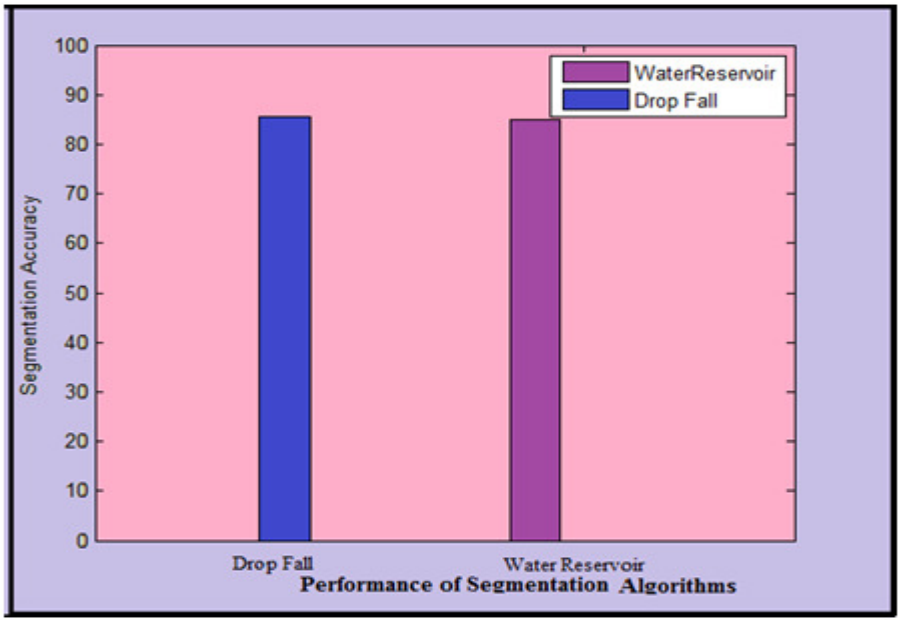


Figure 17 . Segmentation Rate of Drop fall and Water Reservoir techniques

7. CONCLUSION

The system showed good results when tested on the 150 samples of varying degraded epigraphs. It provides better enhanced output on applying filters of appropriate mask size - 4x4 mask size for Median filter, 2x2 mask size for Gaussian blur, 4x4 mask size for Mean and Bilateral filter. Segmentation is carried out using Drop Fall and Water Reservoir algorithms and system can efficiently segment characters from ancient document images. System segments the compound characters correctly when connectivity present. Few cases where in connectivity is absent, compound character is segmented separately. Segmentation rate of 85%-90% for Drop fall algorithm and 85%-90% for Water Reservoir algorithm is achieved.

REFERENCES

[1] Tanzila Saba, Ghazali Sulong & Amjad Rehman, "A survey on Methods and Strategies on Touched Characters Segmentation", International Journal of Research and Reviews in Computer Science (IJRRCS), Vol.1, No.2, June 2010.

[2] M. Trentacoste et al. "Unsharp Masking, Counter shading and Halos: Enhancements or Artifacts", 2012, The Euro graphics Association and Blackwell publishing Ltd. 2012.

[3] Maya R. Gupta, National P. Jacobson, Eric K. Garcia, "Binarization and Image Preprocessing for searching Historical Documents", University of Washington, Seattle, Washington 98195, April 2006.

[4] Bolan Su, Shijian Lu, Chew Lim Tan, "Binarization of Historical Document images Using the local Maximum and Minimum", pp.9-11, June 2010.

[5] N. Venkata Rao, A.V.Srinivasa Rao, S Balaji and L. Pratap Reddy, "Cleaning of Ancient Document Images Using Modified Iterative Global Threshold" ,IJCSI international Journal of Computer Science Issues, Vol.8.Issue 6, No 2, pp. ISSN (online):1694-0814, November 2011.

[6] Dr. B.P.Mallikarjunaswamy, Karunakara K, "Graph Based Approach for Background and Segmentation of the Image", Vol 02, pp.ISSN: 2230-8563, June 2011.

[7] Mamatha H.R and Srikanta Murthy K, "Morphological Operations and Projection Profiles based Segmentation of handwritten Kannada Document", International Journal of Applied Information System (IJAIS)-ISSN: 2249-0868 Foundation of computer Science FCS, New York, USA Vol 4-N0.5, pp. 13-19, October 2012.

[8] Das, Reddy, Govardhan, Saikrishna, "Segmentation of Overlapping text lines, Characters in printed Telugu text document images", International Journal of Engineering science and technology, Vol.2(11), pp.6606-6610, 2010.

[9] Karthik S, Mamatha H.R, Srikanta Murth y K, "An Approach based on Run Length Count for denoising the Kannada Characters", Internatinal journal of Computer applications, Vol 50-No.18, pp.0975-8887, July 2012.

[10] N. Anupama, Ch. Rupa & Prof. E. Sreenivasa Reddy, "Character Segmentation for Telugu Image Document using Multiple Histogram Projections", Global Journal of Computer Science and Technology Graphics & Vision, Vol 13, 2013.

[11] Srinivasa Rao A.V, segmentation of ancient Telugu Text Documents, published online in MECS ,I.J. Image, "Graphics and Signal Processing", pp.8-14 Published Online July 2012.

[12] Partha Pratim Roy, Umapada Pal, Josep Llados, Mathieu Delalandre, "Multi-oriented and Mutisized Touching Character Segmentation using Dynamic programming", 10th international conference on Document analysis and Recognition, DOI 10.1109/ICDAR.124 IEEE 2009.

[13] Indu Sreedevi, Rishi Pandey, N. Jayanthi, Geetanjali Bhola, Santanu Chaudhury2, "NGFICA Based Digitization of Historic Inscription Images, ISRN Signal Processing", Vol 3, 2013.

[14] Pavlidis, T. "Algorithms for Graphics and Image Processing", Computer Science Press, Rockville, Maryland, 2009.

[15] Zhang, T. Y. and Suen, C. Y. "A Fast parallel algorithm for thinning digital patterns. Communication. ACM", Vol 3, No 27, pp 236-239, 2009.

[16] B Gangamma,Srikanta Murthy K,Arun Vikas Singh, "Restoration of Degraded Historical Document Image", Journal of Emerging Trends in computing and Information Science, Vol 3, No 5, May 2012.

[17] Xiaoqing Lu, Zhi Tang, Yan Liu, Liangcai. "Stroke –based character segmentation of low –quality images on ancient Chinese tablet", 12th International conference on document Analysis and recognition, 2013.

[18] U. Pal, A. Belaid and Ch. Choisy "Touching numeral segmentation using water reservoir concept", Pattern Recognition Letters, Vol.24, pp. 262-272, 2003.

[19] S.Khan."Character Segmentation Heurisrtics for Check Amount Verification "Master Thesis, Massachusetts Institute of Technology, 1998.

[20] X.Wang, K. Zheng and J.Guo. "Inertial and Big Drop Fall Algorithm", International Journal of information Technology, vol.12, No.4, 2006.

[21] Parminder Singh and Harjinder Singh, " A Comparison of High Pass Spatial Filters using Measurements and Automation", Int. J. IJERT, Vol 1, May 2012.

[22] Mamatha H.R, Sonali Madireddi, Srikanta Murthy K, "Performance Analysis of various filters for De-noising of Handwritten Kannada Documents", International Journal of Computer Applications (0975 – 888) Volume 48– No.12, June 2012.

[23] Messaoud, I.B, et.al. "Region Based Local Binarization approach for Handwritten Ancient Documents, "in Int. Conf. Front. Handwrit recognit. Bari, pp.633-638, 2012.

[24] Srinivasa Rao A V, D R Sandeep, V B Sandeep,S Dhanam Jaya "Segmentation of Touching Hand written Telugu Characters by using Drop Fall Algorithm, International Journal of Computers & Technology, Volume 3 No. 2, OCT, 2012.

[25] Praveen kumar. C, Kiran. Y. C, " Kannada Handwritten Character Segmentation using Water Reservoir Method" , International Journal of Systems, Algorithms & Application 2012.

INTENTIONAL BLANK

SINGLE IMAGE FOG REMOVAL BASED ON FUSION STRATEGY

V. Thulasika and A. Ramanan

Department of Computer Science, Faculty of Science,
University of Jaffna, Sri Lanka
v.thula.sika@gmail.com, a.ramanan@jfn.ac.lk

ABSTRACT

Images of outdoor scenes are degraded by absorption and scattering by the suspended particles and water droplets in the atmosphere. The light coming from a scene towards the camera is attenuated by fog and is blended with the airlight which adds more whiteness into the scene. Fog removal is highly desired in computer vision applications. Removing fog from images can significantly increase the visibility of the scene and is more visually pleasing. In this paper, we propose a method that can handle both homogeneous and heterogeneous fog which has been tested on several types of synthetic and real images. We formulate the restoration problem based on fusion strategy that combines two derived images from a single foggy image. One of the images is derived using contrast based method while the other is derived using statistical based approach. These derived images are then weighted by a specific weight map to restore the image. We have performed a qualitative and quantitative evaluation on 60 images. We use the mean square error and peak signal-to-noise ratio as the performance metrics to compare our technique with the state-of-the-art algorithms. The proposed technique is simple and shows comparable or even slightly better results with the state-of-the-art algorithms used for defogging a single image.

KEYWORDS

Airlight, Dark channel prior, Direct Attenuation, Fog removal, Image restoration

1. INTRODUCTION

Images of natural scenes are degraded due to bad weather such as fog, haze, mist, rain, smoke etc. The natural phenomena such as fog, smoke occurs mainly due to atmospheric absorption and scattering. While taking the image during bad weather condition, the radiance flux received by the camera from the scene point is attenuated along the line of sight. The incoming light is mixed with the light coming from all other directions called the airlight. In this phenomenon, the amount of scattering depends upon the distance of the scene points from the camera. Therefore there is a significant decay in the colour and the contrast of the captured image.

Fog removal is a fundamental requirement in the field of computer vision applications such as surveillance, remote sensing systems, outdoor object recognition and various camera-based intelligent driver assistance systems. Removal of fog from the input foggy image can

David C. Wyld et al. (Eds) : ACITY, DPPR, VLSI, WiMNET, AIAA, CNDC - 2015
pp. 115–123, 2015. © CS & IT-CSCP 2015 DOI : 10.5121/csit.2015.51310

exponentially increase the visibility of the scene. Recently, many computer vision algorithms suffer from low-contrast scene radiance. In many of the automatic systems, we assume that the input images have clear visibility. But it is not applicable in many of the real world situations. Fog is purely dependent upon unknown depth. Earlier fog removal techniques required multiple images of same scene under different environmental conditions. Such methods cannot be used for dynamic scenes and cannot be used on existing image sets. Therefore many researchers follow a new strategy, which is based on single gray-level or colour image without using any other extra source information. In this regard, several authors have proposed various defogging techniques such as dark channel prior [1], improved dark channel prior with different filters [4,9,10], anisotropic diffusion [8] and Tarel's method [6,7]. Assessing the overall performance of the individual components in such systems is difficult, since the computational requirements and the fine tuning of the different parts become crucial. However, Tarel's method [7] is promising when applied to road images by taking into account that a large part of the image can be assumed to be a planar road whereas the dark channel prior method [1] is not dedicated to road images and thus the road part of the image is over enhanced. In this paper, we propose a technique that combines two derived images from a single foggy image. One of the images is derived using a contrast based method while the other is derived using a statistical based approach. These derived images are then weighted by a specific weight map to restore the image.

The remainder of this paper is organized as follows: Section 2 provides the mathematical model of a foggy image. Section 3 summarises the contrast and statistical based algorithms that have been widely used in the recent years to remove fog from single images. Section 4 explains the proposed fusion strategy in achieving fog removal. Section 5 describes the experimental setup and testing results which supports our claim. Finally, section 6 concludes this paper.

2. BACKGROUND

Fog is a combination of two components: Airlight and direct attenuation. Airlight adds whiteness into scene whereas the attenuation decreases the contrast in the scene as well as variation of scene colour which finally leads to a poor visual perception of the image. In computer vision the model widely used to describe the formation of a fog image can be expressed by the following equation:

$$\text{Foggy Image} = \text{Direct Attenuation} + \text{Airlight} \quad (1)$$

The direct attenuation describes the scene radiance and its decay in the medium. It is a multiplicative distortion of the scene radiance.

$$\text{Direct Attenuation} = J(x) \bullet t(x) \quad (2)$$

where $J(x)$ is the scene radiance and $t(x)$ is the medium transmission.

Airlight or atmospheric veil is caused due to scattering of light. Airlight is an additive one and is a function of the distance between camera and object.

$$\text{Airlight} = A(1 - t(x)) \quad (3)$$

where A is the global atmospheric light.

When the atmosphere is homogenous, the transmission t(x) can be expressed as:

$$t(x) = e^{-\beta d(x)} \quad (4)$$

where β denotes the extinction coefficient of the atmosphere. It indicates that the scene radiance is attenuated exponentially with the scene depth d. Using equations (1), (2), (3) and (4) the intensity value I of *x*th pixel a foggy image can be defined mathematically by equation (5):

$$I(x) = J(x) \bullet t(x) + A(1 - t(x)) \quad (5)$$

The atmospheric light A is estimated as the maximum value of the corresponding region in the input (i.e., foggy) image.

3. RELATED WORK

Tarel and Hautiere [6] proposed a method for fast visibility restoration of images using a median filter algorithm. The restoration works with low complexity for gray and colour images, which adopts white balance, gamma correction, and tone mapping to maintain colour fidelity. Airlight is considered as a percentage between local standard deviation and local mean of the whiteness. Depth map is used to smooth along the corners. This method depends upon linear operations that require many parameters for the adjustment. Their method is not dedicated to road images and thus the road part of the image which is gray is over-enhanced due to the ambiguity between light coloured objects and the presence of fog. Furthermore, in some small edge regions, the desirable defogging results cannot be achieved.

In [7], an extended algorithm of [6] is proposed for better defogging of the roadway area. The method handles road images by taking into account that a large part of the image can be assumed to be a planar road. Thus the extended algorithm introduces a planar constraint to method [6]. The advantages of the improved version are its speed and small number of parameters. Authors claim that the algorithm produces similar quality results with homogeneous fog and it is able to better deal with the presence of heterogeneous fog.

He et al [1] proposed a method based upon dark channel prior which is basically used for single image defogging method. This dark channel prior is mainly used to measure the statistics of the outdoor fog-free image. The authors' method is based on the assumption that some pixels are having very low intensity in any one of the colour channel in the case of regions which do not cover the sky. These pixels are known as the dark pixels. In the case of foggy images, the intensity of the dark pixels is mainly contributed by the airlight. These dark pixels are used to estimate the fog transmission. The aim of the technique is to restore fog free image from the transmission map. Their method employs a dark channel prior which assumes every local patch (15×15) in the fog-free image have at least one colour component near zero. This assumption is sometime violated when there is no black body in some local patches. Instead of using a Markov random field, a soft matting algorithm is used to refine the transmission values. However, refining the observed transmission map with soft matting is computationally expensive. He et al [2] proposed a further refinement to the dark channel prior using a guided image filter in order to reduce the time complexity in [1]. Even though this method greatly reduces the time complexity, the original foggy image chosen as the reference image may lead to incomplete fog removal.

However, the method in [7] is promising when applied to road images whereas the method in [1] is promising when applied to outdoor scene images such as cityscape and forests. Thus we propose a method for a single input image that can handle both road images and outdoor scene images by fusing the methods in [1] and [7] using an appropriate scaling factor. The proposed technique can handle both homogeneous and heterogeneous fog.

4. METHODOLOGY

We formulate the restoration problem based on fusion strategy that combines two derived images, *C* and *S*, from a single foggy image, I. Image *C* is derived using contrast based method proposed in [7] while *S* is derived using statistical based approach proposed in [1]. The overall framework of the fusion strategy is depicted in Figure 1.

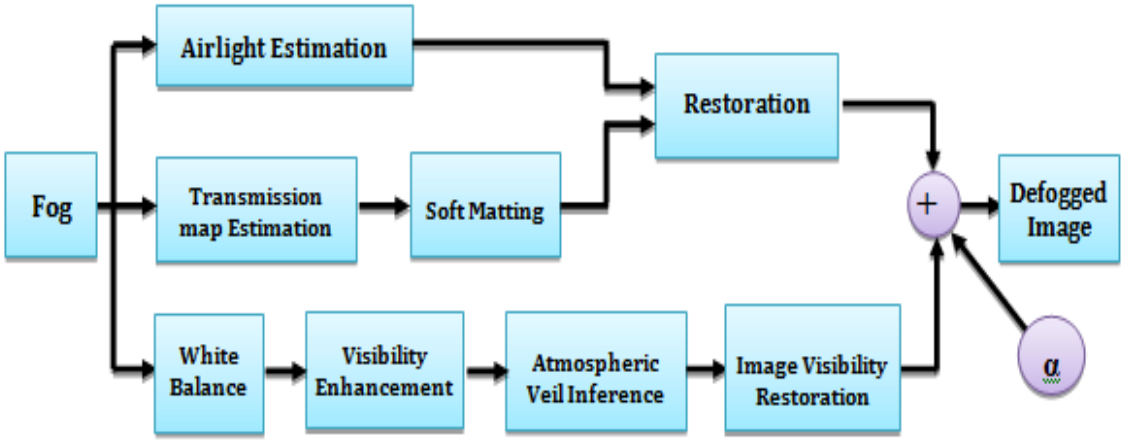


Fig 1: Flow diagram of the proposed fusion based single image defogging technique. The input images for the fusion strategy are obtained by following a statistical based approach on the foggy image as shown in the upper part of the flow diagram and in parallel by using a contrast based approach as shown in the lower part of the diagram. The obtained images are then weighted by a positive scaling factor α to produce defogged image.

4.1. Statistical based approach

To obtain the first input image *S*, a dark channel prior estimation is performed on the original image. It indicates that most local patches in fog-free outdoor images contain some pixels which have very low intensities in at least one colour channel. The atmospheric light is estimated from the most haze-opaque pixel. For example, the pixel with highest intensity is used as the atmospheric light [3]. The atmospheric light A is estimated by the average of the pixels in the original image that correspond to the top lightest 0.1% in the dark channel. Thus, A has three elements, the average values for each colour channel. Following the calculation of A, and assuming that the transmission in a local patch is constant, the transmission can be then estimated. For a refinement purpose, soft matting is applied to the initial transmission. Once we have the transmission map *t* and the atmospheric light A, we can use (5) to recover the scene radiance *S*.

4.2. Contrast based approach

In general, statistical based methods are not dedicated to road images and thus the roadway area in the defogged image is usually over-contrasted. The road can be reasonably assumed to be approximately flat. Thus in [7] the proposed method introduces a planar constraint to the method in [6] for better defogging of the roadway area. To obtain the second input image *C*, the first output is obtained by performing white balance operation on the original image. White balance is performed to correct the colour of the airlight prior to visibility restoration. When the white balance is correctly performed, the fog being pure white, this implies that A in (5) can be set to (1,1,1), also assuming that the input image I is normalised between 0 and 1. Following the white balance operation, a visibility enhancement algorithm is applied in order to avoid keeping certain amount of fog around outliers. The veil is inferred as a percentage of the difference between the local average and standard deviation of the whiteness within the observed image. This is achieved by using a median filter. The classical median filter preserves edges but not corners. This may induce artifacts on very structured scenes such as cityscapes and buildings. In order to preserve edges as well as corners with obtuse angle a median of median along lines filter is used [6]. The restoration of *C* is performed by using the inferred atmospheric veil.

4.3. Fusion strategy

The images *C* and *S* obtained from the contrast based and statistical based methods respectively are then weighted by the following weight map:

$$I = \frac{S + \alpha C}{1 + \alpha}$$

where α is a positive scalar.

Our approach to setting α is to take a subset of the synthetic images, compute the mean square error (MSE) between the fog-free and foggy images and set the scalar to the average sum of values of $\lceil 10\sqrt{MSE} \rceil$. In general α = 2 yields better performance in defogging a single image. The proposed fusion strategy aims to preserve the regions with good visibility.

5. EXPERIMENTAL SETUP

5.1. Dataset

In general it is difficult to obtain the same background-foreground image with and without fog. In this work we make use of two image sets: synthetic and real images. We use images from the Foggy Road Image DAtabase (FRIDA) [7] for synthetic images and images from [1,5] for real images. The database in [7] consists of a total number of 66 colour images of different scenes such as city, highway and rural areas. Each image is of size 640×480. In this database four different types of fog: Uniform, variable sky intensity, variable fog density, and variable sky intensity and fog density were added to images. We have also tested real images from [1,5] that are of cityscapes, buildings and forests. Each image is of size 600×400. A subset of images of road images and real images is shown in Figure 2(a).

5.2. Evaluation criteria

In order to check the robustness of our proposed fusion based defogging approach the mean square error (MSE) and peak signal-to-noise ratio (PSNR) are estimated.

1) *MSE:* The mean squared error represents the cumulative squared error between the compressed and the original image. The lower the value of MSE, the lower the error. Given a noise-free $m \times n$ gray-level image I and its noisy approximation K, MSE is defined as:

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i,j) - K(i,j)]^2$$

2) *PSNR:* The Peak signal-to-noise ratio is computed in decibels between two images. Higher the PSNR, the better the quality of the compressed or reconstructed image. The PSNR is defined as:

$$PSNR = 10.\log_{10}\left(\frac{MAX_I^2}{MSE}\right)$$

$$= 20.\log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right)$$

$$= 20.\log_{10}(MAX_I) - 10.\log_{10}(MSE)$$

where MAX_I is the max intensity value in an image i.e., 1 in double.

5.3. Testing Results

We have performed a quantitative evaluation on 60 images taken from the database in [7] and few real images from [1,5] in which a qualitative analysis is performed. The rest of the six images from the database in [7] are used for estimating the novel parameter α of our approach. The α was varied from 1 to 6 by step size one. In each case of α, MSE and PSNR were measured. In general our empirical results show that α = 2 yields better performance in defogging a single image. Therefore the results we report in Tables 1 and 2 are of α = 2.

Table 1 shows the quantized analysis of the mean square error for a subset of images and the overall MSE in the last row. As mean square error needs to be reduced the proposed algorithm shows better results than the available methods.

As PSNR need to be maximized the main goal is to increase the PSNR as much as possible. Table 2 clearly shows that PSNR is maximum for our proposed method. The method proposed in [7] clearly outperforms the method in [1] when considering MSE and PSNR as shown in the above tables. It has been observed that the proposed approach is comparable or slightly better than other two single image based technique in [1] and [7]. Figure 2 shows the qualitative comparison of the proposed method with methods in [1] and [7]. The restored images using our method are more natural and pleasing

Table 1. COMPARISON OF MSE

Image	Contrast based method [7]	Statistical based method [1]	Fusion method [Ours]
1	0.0114	0.0117	0.0112
⋮	⋮	⋮	⋮
10	0.0097	0.0098	0.0095
11	0.0222	0.0226	0.0201
⋮	⋮	⋮	⋮
20	0.0134	0.0138	0.0131
21	0.0175	0.0188	0.0173
⋮	⋮	⋮	⋮
30	0.0206	0.0207	0.0204
31	0.0162	0.0181	0.0161
⋮	⋮	⋮	⋮
40	0.0148	0.0157	0.0145
41	0.0199	0.0194	0.0192
⋮	⋮	⋮	⋮
50	0.0217	0.0221	0.0213
51	0.0218	0.0221	0.0215
⋮	⋮	⋮	⋮
60	0.0203	0.0178	0.0192
Average	0.0171	0.0182	**0.0168**

Table 2. COMPARISON OF PSNR

Image	Contrast based method [7]	Statistical based method [1]	Fusion method [Ours]
1	67.5509	67.4545	67.6467
⋮	⋮	⋮	⋮
10	68.2442	68.2119	68.3646
11	64.6625	64.5966	65.1051
⋮	⋮	⋮	⋮
20	66.3743	66.2571	66.51
21	66.3174	66.0871	66.366
⋮	⋮	⋮	⋮
30	64.9967	64.9789	65.0451
31	65.4234	65.5383	65.5417
⋮	⋮	⋮	⋮
40	66.416	66.1842	66.5158
41	67.0691	66.3794	67.2119
⋮	⋮	⋮	⋮
50	64.7735	64.6943	64.8412
51	64.7442	64.6869	64.8161
⋮	⋮	⋮	⋮
60	66.6173	66.0919	66.6434
Average	66.1385	65.7917	**66.2123**

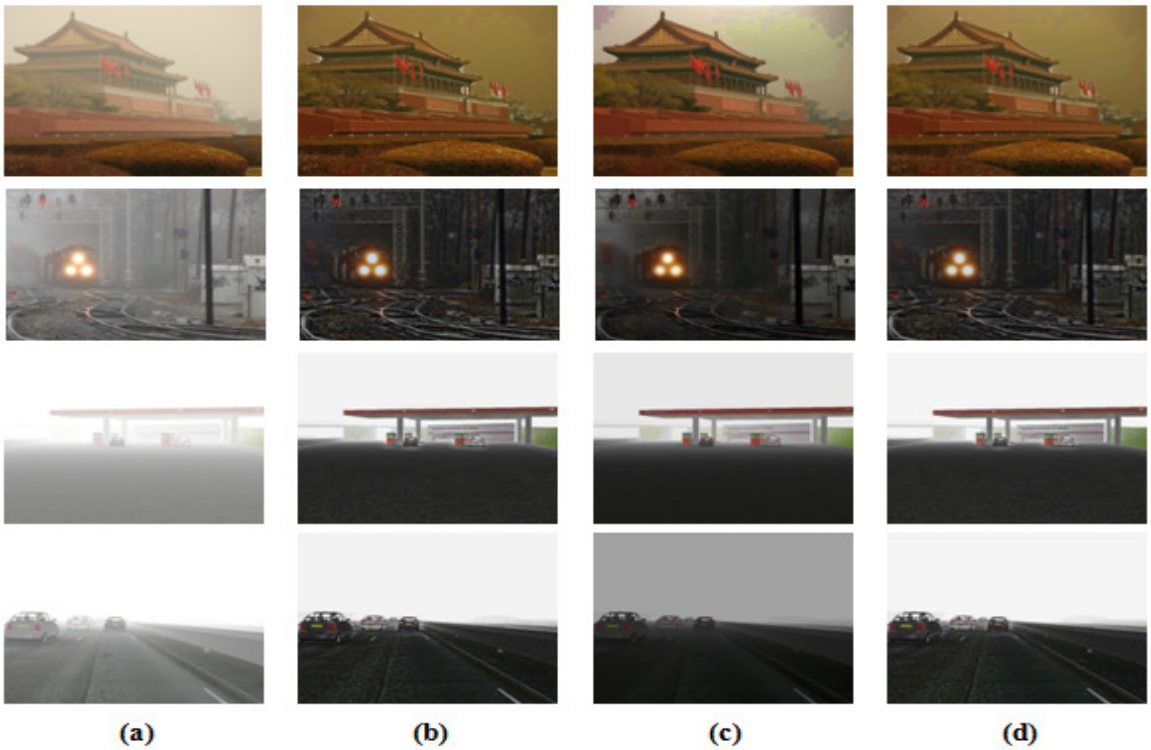


Fig. 2: Comparison of contrast and statistical based techniques with our fusion based strategy. (a) Input image: First two rows are of real images whereas the last two rows are of road images (b) Tarel's results (c) He's results (d) Our results.

6. CONCLUSION

In this paper, a simple but efficient method is proposed to improve single image defogging based on a fusion strategy, which can work well for synthetic (i.e., road images) and real images. The fusion based defogging approach can effectively restore image colour balance and remove fog. This technique is based on selection of appropriate weight map to fuse contrast and statistical based approaches in single image defogging applications. Moreover, it has been observed that this approach outperform the other single image based defogging techniques. The restored images are more natural and pleasing. It has been proved that our method has strong robustness and high availability which can be widely applied to colour images. The proposed method is faster and yields accurate results.

REFERENCES

[1] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior", In proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, pp. 1956–1963, 2009.

[2] K. He, J. Sun and X. Tang, "Guided image filtering", In proceedings of the Eleventh IEEE European Conference on Computer Vision, pp. 1–14, 2010.

[3] S.G. Narasimhan and S.K. Nayar, "Contrast restoration of weather degraded images", In IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, pp. 713–724, 2003.

[4] J. Pang, A. Oscar, and G. Zheng, "Improved single image dehazing using guided filter", In proceedings of the APSIPA Annual Summit and Conference, ACM, pp.1–4, 2011.

[5] Real images: https://sites.google.com/site/computervisionadinastoica/final-project

[6] J. P. Tarel and N. Hautiere, "Fast visibility restoration from a single color or gray level image", In proceedings of IEEE International Conference on Computer Vision, pp. 2201–2208, 2009.

[7] J.P. Tarel, N. Hautiere, A. Cord, D. Gruyer, and H. Halmaoui, "Improved visibility of road scene images under heterogeneous fog", In proceedings of IEEE Intelligent Vehicle Symposium, pp. 478–485, 2010.

[8] A.K. Tripathi and S. Mukhopadhyay, "Single image fog removal using anisotropic diffusion", In proceedings of IET Image Process, vol. 6, no. 7, pp. 966–975, 2012.

[9] S. Yanjuan, R. Liu and W. He, "Image Haze Removal of Wiener Filtering Based on Dark Channel Prior", In proceedings of Eighth IEEE International Conference on Computational Intelligence and Security, pp. 318–322, 2012.

[10] Yan Wang and Bo Wu, "Improved Single Image Dehazing using Dark Channel Prior", In proceedings of IEEE Conference on Intelligent Computing and Intelligent Systems, vol. 2, pp. 789–792, 2010.

AUTHORS

V. Thulasika is an Associate Software Engineer at WSO2 Solutions, Sri Lanka. She received her BSc Honours in Computer Science (2015) from the University of Jaffna, Sri Lanka. Her research interest includes computational photography and image processing.

A. Ramanan is a Senior Lecturer at the Department of Computer Science, University of Jaffna. He received his BSc Honours in Computer Science (2002) from the University of Jaffna, Sri Lanka and his PhD from the University of Southampton, UK (2010). His research interests include computer vision and machine learning.

INTENTIONAL BLANK

IMPLEMENTATION OF VEDIC MULTIPLIER USING REVERSIBLE GATES

P. Koti Lakshmi[1], B Santhosh Kumar[2], Prof.Rameshwar Rao[3]

[1]Assistant Professor,Department of ECE, UCE,Osmania University, Hyderabad.
lakshmi_ponnuri@yahoo.com
[2]Student, ME(ESVLSID), Department ofECE,UCE,Osmania University, Hyderabad santhosh_budati@yahoo.com
[3]Professor(Retd) , Dept of ECE,UCE,Osmania University, Hyderabad.
rameshwar_rao@hotmail.com

ABSTRACT

With DSP applications evolving continuously, there is continuous need for improved multipliers which are faster and power efficient. Reversible logic is a new and promising field which addresses the problem of power dissipation. It has been shown to consume zero power theoretically. Vedic mathematics techniques have always proven to be fast and efficient for solving various problems. Therefore, in this paper we implement Urdhva Tiryagbhyam algorithm using reversible logic thereby addressing two important issues – speed and power consumption of implementation of multipliers. In this work, the design of 4x4 Vedic multiplier is optimized by reducing the number of logic gates, constant inputs, and garbage outputs. This multiplier can find its application in various fields like convolution, filter applications, cryptography, and communication.

KEYWORDS

Multipliers, Urdhva Tiryagbhyam algorithm, Reversible Logic, Vedic Multiplier, Optimization, Quantum cost.

1. INTRODUCTION

A digital signal processor (DSP) is an integrated circuit designed for high-speed data manipulations, and is used in audio, communications, image manipulation, and other data-acquisition and data-control applications. The arithmetic operations performed by most of the DSPs are addition, subtraction which are simple and multiplication, division are complex. The simple multiplication operation may consume many cycles to complete the operation. This causes the processor to become quite slow. To overcome this problem UT multiplier is used. The main constraints of any embedded system are low Power dissipation, high Speed, less Area. The speed of the processor can be increased by using the Vedic Mathematics . The minimum power dissipation is one of the main requirements of the system. The power dissipated in the system can be reduced by introducing Reversible logic. The power dissipation of the reversible logic under idle conditions is Zero. Multiplier is the most chief element in the computing systems such as Digital signal processing, microprocessor, FIR filter etc. So the performance of these application

David C. Wyld et al. (Eds) : ACITY, DPPR, VLSI, WiMNET, AIAA, CNDC - 2015
pp. 125–134, 2015. © CS & IT-CSCP 2015 DOI : 10.5121/csit.2015.51311

can be improved by optimizing the various parameter of the multiplier such as power, speed, area and fault tolerance property. Since these parameter are very much important for Reversible logic circuit or information lossless circuit has zero internal power dissipation and also there are few families of reversible gate that have inherent fault tolerant. This tolerant property in reversible circuit have application in variety of emerging technology such as quantum computing, nanotechnology etc. According to the Moore's law ,by the 2020 the basic memory components of a computer will the size of the individual atoms. At such scales current theory of computer will be fail and an quantum computing reinvented the theory of computer science, Quantum computer can complete task in the breathtakingly time with no internal power dissipation. Multiplication process involves generation of partial products, addition of partial products and finally total product is obtained. So the performance of the multiplier depends on the number of partial products and the speed of the adder.

Vedic mathematics has 16 formulae for performing arithmetic calculation. An Urdhva Tiryakbahayam formula is used for he multiplication, application for all types for multiplication. its literal means " Vertical and Cross-wise" which enhance the speed of multiplication operation. This paper deals with the survey and comparison of the various multiplier mainly in terms of the power, delay, quantum cost. From the survey it is find that the reversible Vedic multiplier based on the Urdhva tiryagbhyam aphorisms is offer the best results in terms of delay, area, power and quantum cost.

In the Array Multiplier ,generation of partial products and addition of that will take more time. Time is an important factor for any computing system. So, in this paper we are proposing the Vedic Multiplier which give results quicker than array multiplier and also Vedic Multiplier using reversible gates is designed with less TRLIC to decrease the power dissipation. In this paper we are also designing the signed vedic multiplier for multiplication of signed numbers.

Vedic Multiplier using reversible gates was proposed by different authors[2][3][4] and they had calculated TRLIC. In this paper we are designing the optimized Vedic multiplier and comparing with that. In this paper, we are discussing basic reversible gates, algorithm of 2x2 vedic multiplier and 4x4 vedic multiplier and how it is optimized by reducing number of gates, garbage outputs, quantum cost, constant inputs.

2. REVERSIBLE GATES

2.1. Reversible Logic Gates

2.1.1 Feynman Gate*:* It is 2x2 gate [3]. If the first input i.e. A is given as 1 the second output will be the complement of the second input i.e. B. so, this gate is also known as Controlled Not Gate. It can also be used to copy inputs. Quantum cost of this gate is one.

2.1.2 Peres Gate*:* It is a 3x3 gate [3]. It can be used as a half adder with third input i.e. c as 0.It also serves the purpose of fan out. Quantum cost of this gate is four.

2.1.3 HNG Gate*:* It is a 4x4 gate [3]. A single HNG gate can serve as a one bit full adder. Quantum cost of this gate is six.

2.1.4 BVPPG gate: In this 5x5 reversible gate [3] is proposed. It is basically for multiplication and can generate two partial products at a time. Quantum cost of this gate is ten.

The basic reversible gates are shown in Fig 1.

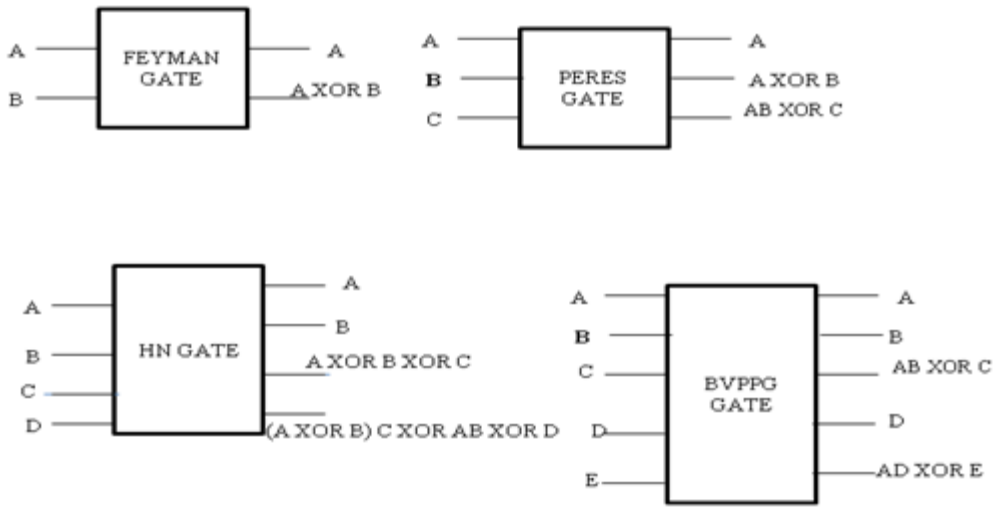


Fig.1 Reversible logic gates

3. MULTIPLICATION USING URDHVA TIRYAGBHYAM SUTRA

A Vedic maths offers two sutras – Urdhva Tiryagbhyam sutra and Nikhilam Sutra for multiplication. Nikhilam sutra is best used for numbers which are nearer to the base of 10,100, 1000 and increased power of 10, whereas Urdhva Tiryagbhyam can be used for any multiplication. The most powerful Vedic multiplication sutra Urdhva Tiryagbhyam means „Vertically and Crosswise". This technique is applicable for any type of number system. General procedure for Urdhva Tiryagbhyam.

Algorithm: Let us consider two digit (for binary number system consider 2 bits) multiplicand and multiplier as "A1 A0" and "B1 B0" respectively and the result as R3R2R1R0.

- Multiplication starts with LSB of the operands i.e. vertical multiplication of A0 and B0 will generate the LSB of the result i. e. R0. For binary numbers no carry will be generated at this stage.

$$R0 = A0B0 \quad \text{..... (3)}$$

- R1 is obtained by crosswise multiplication of A0, B1 and A1, B0 and then adding the two products. In this stage crosswise multiplication and simultaneous addition of the product generates R1 as sum and carry say C1.

$$C1R1 = A0B1 + A1B0 \quad \text{..... (4)}$$

- Again the vertical multiplication between two MSB of the operands i. e. A1 and B1 takes place and product is added with the generated carry C1 in the previous stage to give the third bit of result e. R3 as sum and fourth bit R4 as carry.

$$R3R2 = A1B1 + C1 \quad \text{.....(5)}$$

- Final result is obtained by concatenating R3, R2, R1, and R0. This method is applicable for n number of bits.

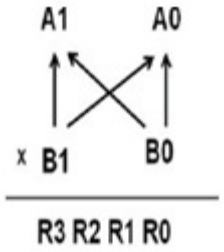


Fig.2 Vertically and Crosswise Multiplication

3.1 2x2 Vedic Multiplier

The 2x2 Vedic multiplier is implemented using 4 equations mentioned below and the logical diagram is shown in fig 3.

q0 = a0.b0 .....(6)
q1= (a1.b0) xor (a0.b1) .....(7)
q2= (a0.a1.b0.b1) xor (a1.b1) .....(8)
q3=a0.a1.b0.b1.....(9)

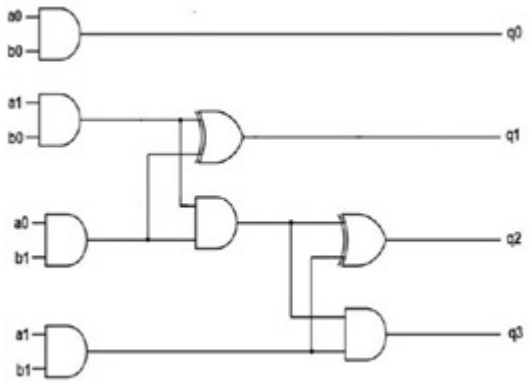


Fig.3 2x2 Binary vedic multiplier

The reversible implementation the circuit uses five Peres gates and one Feynman gate as shown in fig 4. This design has a total quantum cost of 21, number of garbage outputs as 11 and number of constant inputs 4. The gate count is 6. This design does not take into consideration the fan outs. The overall performance of the UT multiplier is scaled up by optimizing each individual unit in terms of quantum cost, garbage outputs etc.

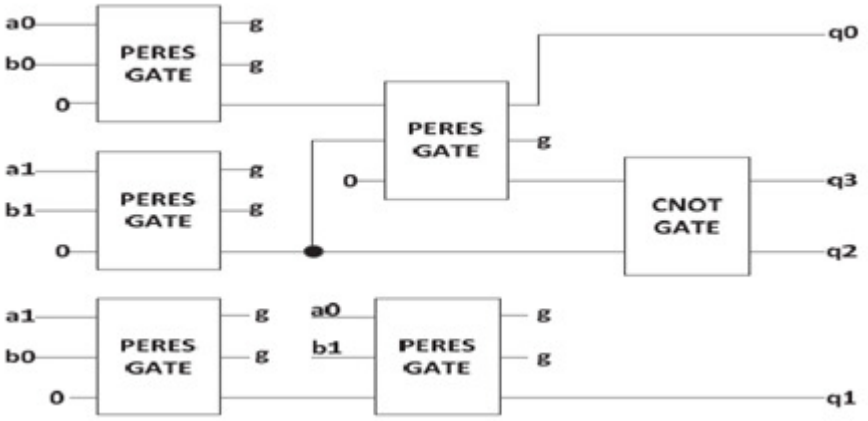


Fig.4 Non optimized 2x2 vedic multiplier

3.1.1 Optimized 2x2 Vedic Multiplier

Reversible implementation is done using a BVPPG gate, three Peres gates and a Feynman gate as shown in fig 5. BVPPG gate generates two partial products among which, one is Q0. Q1 is obtained from one of the Peres gates and Q2, Q3 are the outputs from Feynman gate. This design needs five reversible logic gates, five constant inputs and generates five garbage outputs. Quantum cost and TRLIC of this implementation are 23 and 38 respectively. In this implementation fan out of every signal including primary inputs is one.

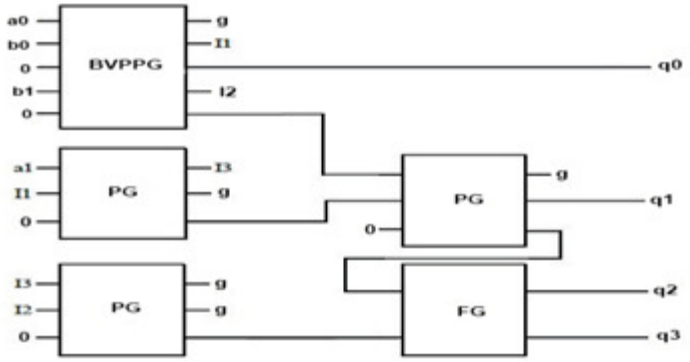


Fig.5 Optimized 2x2 vedic multiplier

3.2 4x4 Vedic Multiplier Implementation

Block diagram of 4x4 is shown in Fig. 6. In this block four 2x2 multipliers are arranged systematically. Each multiplier accepts four input bits; two bits from multiplicand and other two bits from multiplier. Addition of partial products are done using two four bit ripple carry adder and 5bit rca .

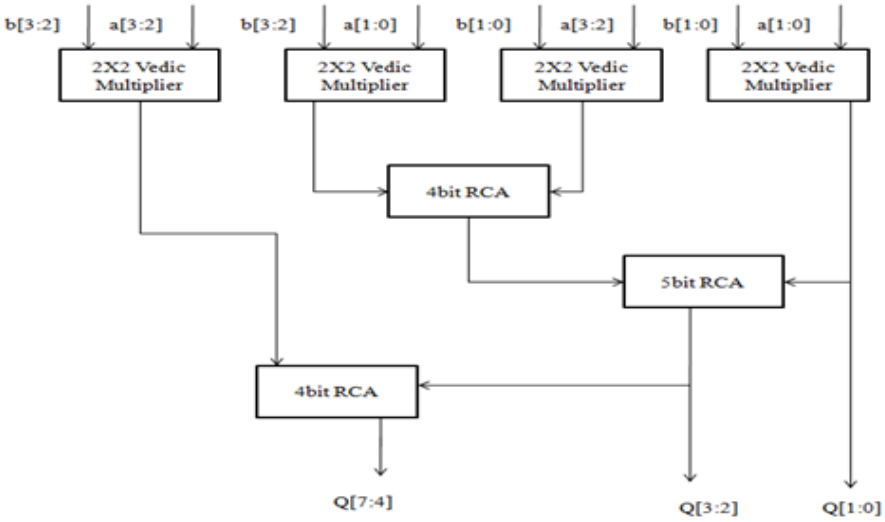


Fig.6 Non optimized 4x4 vedic multiplier

3.2.1 Optimized 4x4 Vedic multiplier

Block diagram of 4x4 is shown in Fig. 7. In this block four 2x2 multipliers are arranged systematically. Each multiplier accepts four input bits; two bits from multiplicand and other two bits from multiplier. Addition of partial products are done using two four bit ripple carry adder, a two bit ripple carry adder and a half adder. We obtain the final result by concatenating the last two bits of the first multiplier, four sum bits of the second four bit ripple carry adder and the sum bits of two bit ripple carry adder.

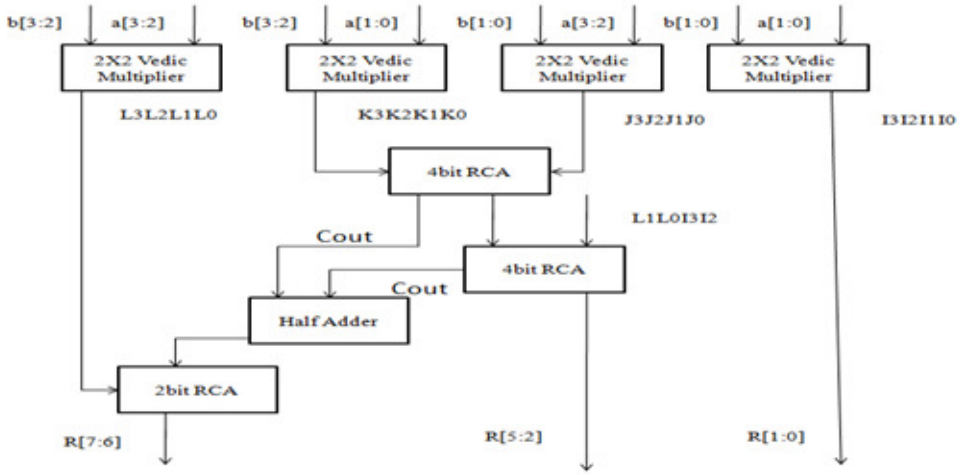


Fig.7 Optimized 4x4 Vedic multiplier

The comparison of optimized and non optimized vedic multiplier are shown below in Table No.1. by considering the parameters of teversible gates like number of gates, constant inputs, garbage outputs and quantum cost. In total we will add this all parameters to find out the total reversible logic implementation cost[TRLIC].

Table 1. Comparison of 4x4 vedic multiplier

Multiplier	No of gates	Constant inputs	Garbage caount	Quantum cost	TRLIC
Non Optimized 4x4 vedic Multiplier using non optimized 2x2 mutiplier	37	29	62	162	290
Non Optimized 4x4 vedic Multiplier using optimized 2x2 mutiplier	33	33	43	164	273
Optimized 4x4 vedic Multiplier using non optimized 2x2 mutiplier	37	27	52	148	264
Optimized 4x4 vedic Multiplier using optimized 2x2 mutiplier	31	31	40	156	258

From the comparison Table No.1 we can see that our design requires less number of gates compare to other multipliers. Garbage outputs and quantum cost is also less. Constant inputs required is lesser than four other multipliers. Significant reduction in quantum cost and TRLIC is observed. So, we can say our design is optimized as compare to other designs exist in terms of number of gates, constant inputs, garbage outputs, quantum cost, and TRLIC.

5. SIGNED VEDIC MULTIPLIER

5.1. Algorithm for signed Multiplier

Step1: First we are declaring inputs and outputs.

Step2: Now we are taking MSB bits of both inputs which is a sign bit and now we are calculating XOR of both the bits which indicates sign of the result.

Step 3: Now the negative numbers which are in 2's complement form should convert to original Form for this we are subtracting the number with 4 in the case of 2x2 and 16 in the case of 4x4 if at all MSB bit of number is 1.

Step 4: After converting the numbers we are going to call Reversible Unsigned Vedic Multiplier We get the multiplier output.

Step 5 : After getting output from the multiplier we are again converting the number in to 2's Complement form if and only if output XOR output of MSB's of input is 1 otherwise We are taking the direct output.

Signed Vedic Multiplier is used to add signed numbers. Usually in our system negative numbers will be represented in 2's complement form. So when we are declaring inputs they will be in 2's complement form. The usage of unsigned vedic multiplier function is good for normal binary form. To use that function we have to convert 2's complement to normal form and we can call undigned vedic multiplier. The conversion can be made by subtracting the number excluding the sign bit with the corresponding 2^n like for 2bit number we have to subtract with 4, for 4bit we have to subtract with 16.

Now we are taking xor of two MSB bits to get the output sign. For example two MSB bits are 1 and 0 the output sign consists of 1 which represents that result output is negative number. After

converting the negative number to normal form we are going to call corresponding unsigned multiplier and the result will be in normal form.

Now the output result is converted in to normal form if and only if MSB ouput is 1otherwiswe the output will be taken same . In this signed vedic multiplier the most important step is converting negative numbers in to normal form.

6. RESULTS AND COMPARISON

Simulations are carried out using Xilinx 13.1 and synthesized for Spartan3e XS5OO series target board and results are comapared for vedic multilier implemented with normal and reversible gates for the parameters path delay, routing delay and total power(Table No2). The proposed multiplier is also compared with array multiplier for different word sizes as 4bit,8bit, 16 bit and32bit (Table No 3).

Table No. 2 Comparison of 16x16 normal and Reversible vedic multipliers

Parameter	Normal Vedic Multiplier	Vedic Mutiplier Using Reversible gates
Path delay(ns)	60.23	49.101
Logic delay(ns)	32.12	28.426
Routing delay(ns)	28.11	20.625
Dynamic power(mw)	2.35	2
Total power(mw)	52.21	49.21

From Table No.3 we can say that our proposed multiplier i.e. vedic multiplier is faster than array multiplier.

Table No. 3 Comparison of Varous widths of Vedic and Array Multiplier.

	4x4 Multiplier		8x8 Mutiplier		16x16 multiplier		32x32 Mutiplier	
Parameter	Vedic Multiplier	Array Multiplier	Vedic Multiplier	Array Multiplier	Vedic Multiplier	Array Multiplier	Vedic Multiplier	Array Multiplier
Path dela(ns)	15.36	27.36	25.54	47.12	49.101	92.57	87.587	126.33
Logic delay(ns)	10.12	16.12	15.75	28.01	28.426	52.14	50.25	68.171
Routing delay(ns)	5.24	11.24	9.78	19.11	20.675	40.43	37.337	58.164
Dynamic power(mw)	0.5	0.51	1.1	1.12	2	2	3	3
Total power(mw)	12	12.12	26.63	26.7	49.21	49.21	79	79

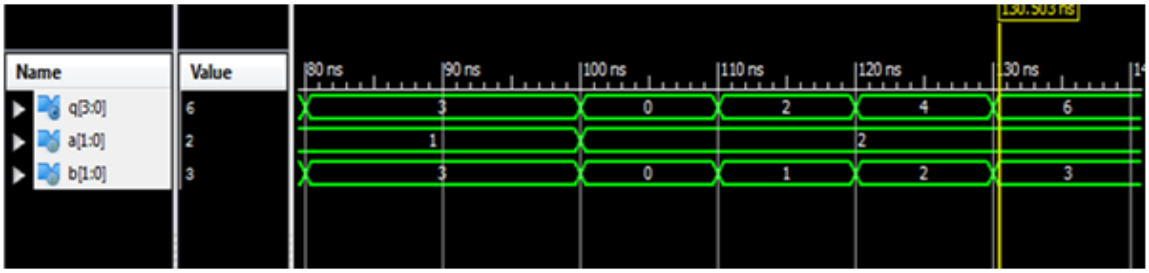


Fig.8 2x2 vedic multiplier

In Fig 8. the simulation result of 2x2 Vedic Multiplier is shown. In this Fig.8 a and b are two inputs of 2bit length and output q of 4bit.

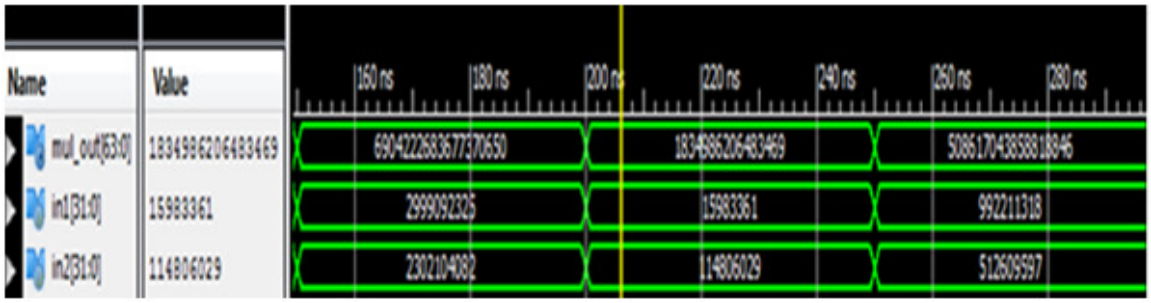


Fig.9 32x32 Vedic Multiplier

Fig.9 shows the simulation result of 8x8 Vedic Multiplierwhere ‘in1’ and ‘in2’ are two inputs of 32bit length and output taken as mul_out of 64bit.

Fig.10 8x8 signed Vedic Multiplier

Fig 10. shows the simulation result of 8x8 Signed Vedic Multiplierwhere ‘a’ and ‘b’ are two inputs of 4bit length and output taken as ‘h’ of 8bit.

8. CONCLUSION

Vedic mathematics is long been known but has not been implemented in the DSP and ADSP processors employing large number of multiplications in calculating the various transforms like FFTs and control applications such as P, PI, PID Controller implementing in FPGA etc. The proposed Vedic multiplier proves to be highly efficient in terms of speed. Due to its regular and parallel structure it can be realized easily on silicon as well. The main advantage is delay increases slowly as input bits increase. Vedic multiplier can be efficiently adopted in designing Fast Fourier Transforms (FFTs) Filters and other applications of DSP like imaging, software defined radios, wireless communications.

REFERENCES

[1] H. R. Bhagyalakshmi, M. K. Venkatesha, “An Improved Design of a Multiplier using Reversible Logic Gates,” IJEST, Vol. 2, No. 8, 2010

[2] Rakshith T R and Rakshith Saligram, Design of High Speed Low Power Multiplier using Reversible logic: a Vedic Mathematical Approach, Intl. Conf. on Circuit, Power and Computational Technologies.

[3] Vijay K Panchal, Vimal H Nayak, "Analysis of multiplier Circuit using Reversible Logic" , International Journal for Innovative Research in Science & Technology.
[4] Prof . Amol D. Morankar, Prof Vivek M.Sakode, " Reversible Multiplier with Peres Gate and Full Adder", International Journal of Electronics Communication and Computer Technology, Volume 4,Issue 4,pp-2249-7838, July 2014.
[5] A. Shifana Parween and S. Murugeswari, " A Design of High Speed, Area Efficient, Low Power Vedic Multiplier using Reversible Logic Gate", International Journal of Emerging Technology and Advanced Engineering, Volume 4,Issue 2, February 2014.
[6] Krishnaveni D and Umarani, " VLSI Implementation of Vedic Multiplier with Reduced Delay", International Journal of Advanced Technology & Engineering Research, Volume 2,Issue 4, July 2012
[7] Swami Bharati Krishna Tirtha, Vedic Mathematics. Delhi: Motilal Banarsidass publishers 1965
[8] Ch. Harish Kumar , "Implementation and Analysis of Power, Area and Delay of Array, Urdhva,Nikhilam Vedic Multipliers", International Journal of Scientific and Research Publications, Volume 3, Issue 1,January 2013.
[9] https://www.cs.princeton.edu/courses/archive/fall04/cos576/papers/bennett73.html
[10] R. Landauer, "Irreversibility and Heat Generation in the Computational Process", IBM Journal of Research and Development, 5, pp. 183-191, 1961.

AUTHORS

Mrs. P. Koti Lakshmi is working as Assistant Professor in the department of ECE at Osmania University , Hyderabad. She has 15 years of teaching experience. She has obtained her AMIETE Degree from IETE, New Delhi in 1999, her M.E in Digital systems in 2004 and persuing Ph.D in VLSI Design from Osmania University , Hyderabad, Andhra Pradesh . Her areas of interest include VLSI Design and Wireless Communications.

B Santhosh Kumar is a student of ME.UCE,OsmaniaUnviversity,Hyderabad. He completed his B Tech from CVR College of Engineering , Hyderabad in the year 2012. His interests are Low power VLSI design circuits, digital circuit designs.

Prof. Rameshwar Rao is the Retired Professor from department of ECE, University college of Engineering,Osmania University,Hyderabad. During his tenure he held many positions as Vice Chancellor of JNTUH, Hyderabad, Andhra Pradesh., Dean, Faculty of Engineering, Osmania University (OU)., Convener, PGECET. He obtained B.E. degree in ECE from OU, M.Tech. and Ph.D. degree from the prestigious IIT Bombay. His work experience spans across 35 years as R&D engineer at Avionics Design Bureau, Hindustan Aeronautics Ltd., Hyderabad and as an eminent teacher at Osmania University,Hyderabad. His reach interests include VHDL Modeling & Synthesis, (During last three years),Data and Computer Communications, Detection and Estimation Theory,Information and Coding Theory, Microprocessor based applications and VLSI Design. He has to his credit more than 60 conference/ journal publications.

HIGH SPEED LOW POWER CMOS DOMINO OR GATE DESIGN IN 16NM TECHNOLOGY

P. Koti Lakshmi[1] and Prof. Rameshwar Rao[2]

[1]Assistant Professor Dept. of ECE, UCE, Osmania University, Hyderabad.
lakshmi_ponnuri@yahoo.com
[2]Professor (Retd), Dept of ECE, UCE, Osmania University, Hyderabad.
rameshwar_rao@hotmail.com

ABSTRACT

Dynamic logic circuits provide more compact designs with faster switching speeds and low power consumption compared with the other CMOS design styles. This paper proposes a wide fan-in circuit with increased switching speed and noise immunity. Speed is achieved by quickly removing the charge on the dynamic node during evaluation phase, compared to the other circuits. The design also offers very less Power Delay Product (PDP). The design is exercised for 20% variation in supply voltage.

KEYWORDS

Low PDP design, High speed OR gate, Domino OR gate, Low power design.

1. INTRODUCTION

The rapid advancements in the field of VLSI is due to the increased use of battery operated devices such as laptops, PDAs, mobiles etc., advancements in wireless communications and computations are the urge for low power budgets and compactness. To achieve this, the transistor size has been continually scaled down and to have proper operation of the device, the supply voltages have also been scaled. As the technology aggressively scales down, the density on the chip has increased and hence the interconnection density, which increased the coupling capacitance of the circuit. This lead to increased interaction between the connections and thereby increasing crosstalk and system failures. On the other hand with the decrease in the supply, the gate threshold is decreased to preserve system throughput and so leakage currents have increased.

Dynamic logic circuits found their wide application in high speed, low power areas such as microprocessors, digital signal processing, dynamic memories etc., because of their low device count, high speed, short circuit power free and glitch free operation [2]. On the other hand it is also possible to design a dynamic logic unit that is smaller than its static counterpart. Dynamic logic consists of pull down network realizing the logic. From the basic theory of dynamic logic the circuit is pre-charged and evaluated at every clock cycle. When a dynamic gate is cascaded by a static inverter, it is called Domino logic. Due to high clock frequency, a large amount of noise

David C. Wyld et al. (Eds) : ACITY, DPPR, VLSI, WiMNET, AIAA, CNDC - 2015
pp. 135–141, 2015. © CS & IT-CSCP 2015 DOI : 10.5121/csit.2015.51312

gets induced and power consumption increases. The main draw backs in dynamic logic are charge sharing and cascading. To overcome these problems domino logic is used. Domino gates runs faster than the static gates as they present much lower input capacitance for the same output current and a lower switching threshold. In this paper we are proposing a technique to reduce the power and increase the speed of domino gate.

1.1 PROBLEM STATEMENT:

The basic domino logic stage consists of logic realized using N-MOS (Mn) in pull down network and the pull up net work consists of a single P-MOS (Mp) to pre charge the dynamic node to logic high as shown in Fig.1. The dynamic node is cascaded into a static inverter from where the gate output is taken and can be connected to the N-FET input of the next stage[1]. When clock =0, the dynamic node charges to V_{dd} and the bottom transistor Mn is responsible for holding the charge on the dynamic node irrespective of the input combination applied to the pull down network. Thus the output goes to logic 0 during this interval (pre- charge phase). When the clock = 1(evaluation phase) the pre-charge transistor (Mp) goes off, allowing the dynamic node to settle down to a state determined by the inputs . Based on the logic implemented, the charge on the dynamic node may be retained at logic 1, thus output remains at logic 0 or the dynamic node may get discharged to logic 0 and output may rise to logic 1.

During evaluation phase when all the inputs are at logic 0, dynamic node should be at logic 1, but the wide fan-in N-MOS leaks the charge stored on the dynamic node due to sub threshold leakage. This is again compensated by P-MOS keeper (Fig.2), which aims to restore the charge on the dynamic node. But when a noise pulse occurs at any of the input such that pull down network provides a direct path to ground, the keeper may not be able to retain the charge on the dynamic node and the node gets wrongly discharged. As the noise in Domino gates is becoming more important than area, power and delay issues in the sub micro meter regime , recently several techniques have been proposed [6],[7] to reduce noise in domino circuits. All the techniques have aimed at reducing the noise effect, but have several drawbacks related to area, power and delay.

In section II existing domino techniques were discussed, section III discusses the proposed scheme, simulation and results compared with other existing schemes is presented in section IV and section V presents the conclusion.

2. BACK GROUND AND RELATED WORK

To compensate the leakage at dynamic node a weak transistor called keeper transistor is used. It prevents the charge loss and keeps the dynamic node at strong high when pull down network is off. In the first Domino proposal [3] the gate of the keeper is connected to ground which makes it always ON. Thus at the beginning of the evaluation phase if the pull down network turns ON, the dynamic node tends to discharge through PDN and keeper starts injecting the lost charge to the dynamic node as it is always ON, which results in contention. This technique introduced a potential DC power consumption. In order to reduce this extra power dissipation, a feedback keeper was proposed in [4],[5]. In this the PMOS gate of the keeper (Fig.2) is connected to the output of the static inverter. Thus during Pre-charge the dynamic node is at high, and the keeper remains on and during evaluate phase if the pull down net work is on, dynamic node gets

discharged and the output node is at logic high which makes the keeper transistor off, thus eliminating contention.

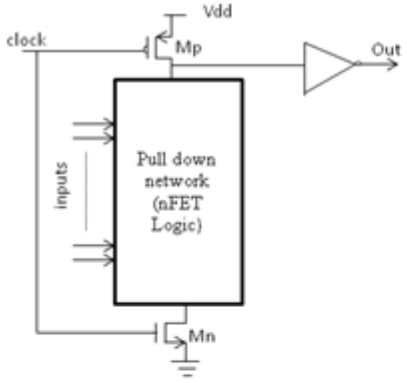


Fig.1 Basic Domino logic stage

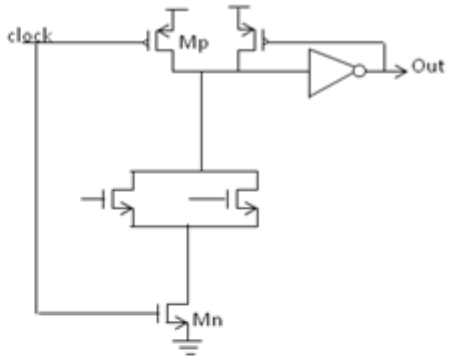


Fig.2 Standard Domino OR gate

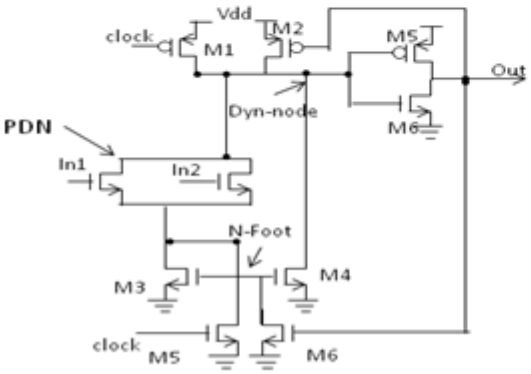


Fig .3 Diode Footed Domino[6]

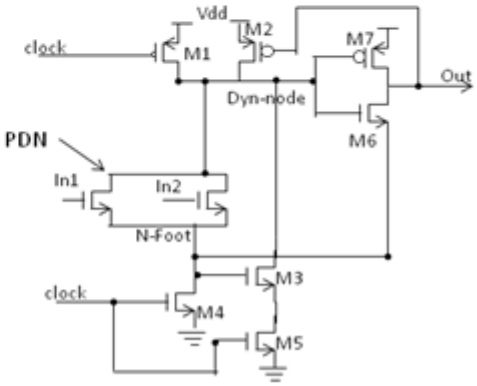


Fig.4 Domino circuit in Scheme [8]

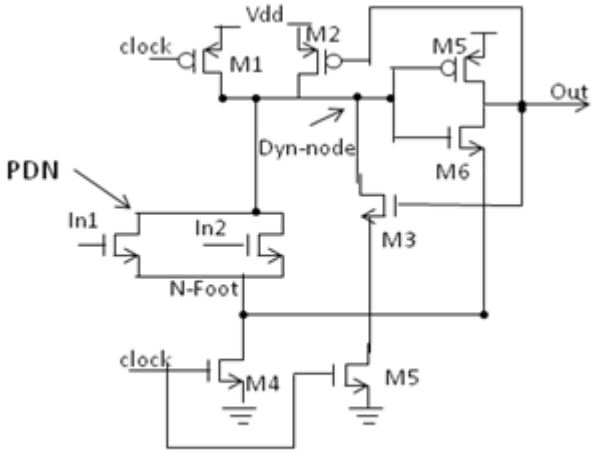


Fig.5 Proposed Domino circuit

In [6] a diode footed Domino was proposed (Fig.3), where an NMOS transistor Mn is connected in diode configuration. In this configuration, the leakage current flowing through the PDN in evaluation phase causes the drop across the diode transistor making Vgs negative, thus reducing leakage current. The performance degradation can be compensated by the mirror network. The inverted clock increases the capacitive load of the clock driver.

In [7] the circuit is based on pull-up network, consisting of only n-MOS transistors. This style does not have a p-MOS transistor. When the clock is low the pre-charge transistor M1 is switched on and the dynamic node is charged to 0v. When the clock is high, M1 is off and the dynamic node gets conditionally charged by the pull-up network to Vdd, but due to the absence of pull-up transistor, the node gets charged to Vdd-Vth and this drop is compensated by M2 the keeper transistor. This circuit needs an inverted clock which increases the capacitive load and area to invert the clock.

In [8] an additional evaluation transistor M5 is added in order to stack M3 and make its gate- to source voltage small, thus making the circuit noise robust and less leakage power consuming (Fig. 4). The performance degradation is compensated by widening the keeper transistor M2.

3. PROPOSED SCHEME

The proposed domino circuit is as shown in the figure 5. Transistor M3 and M5 are connected between the dynamic- node and ground and the gate ofM3 transistor is connected to the OUT terminal and M3 is stacked with M5. During evaluation phase when PDN is on with one or more inputs connected to logic one, the transistor M4 discharges the dynamic-node and the Out terminal goes to logic ‘1’ and M3 becomes on which aids in faster discharge of any accumulated charge on dyn_node along with PDN and M4. The rate of discharge can be controlled by changing the W/L ratio of M4. When all the inputs are at logic ‘0’, output stay at logic ‘0’ and M3 remains off and thus dyn_node retains its charge. If any input changes from logic ‘0’ to logic ‘1’ , PDN becomes conducting and during evaluation phase when clock is high, dyn_node discharges below the threshold voltage of the inverter turning its output to logic ‘1’. When output becomes logic ‘1’, M3 turns on providing a path to discharge dyn_node quickly as M5 is also ON during evaluation phase. At the same time as the source node of M6 being connected to N-foot, effect of noise in the circuit can also be reduced.

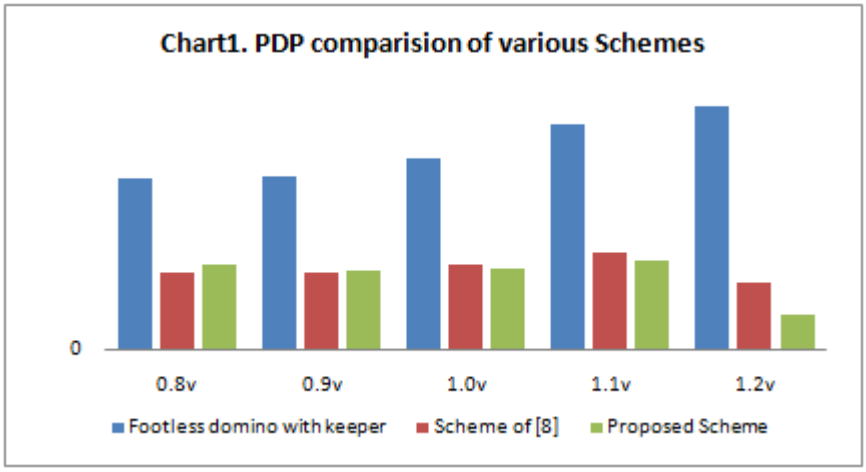
4. SIMULATION AND RESULTS

The circuits were simulated using Tanner T-spice using 16nm technology with 1V supply. The circuit was compared with OR gate of the existing techniques. The OR gate was implemented because it is a typical example of wide pull down network. It is found that the proposed circuit performs better than the previous circuits. The power and delay were measured using T-Spice and PDP was calculated. Delay for various circuits is measured using window technique. Table.1 shows the power and delay measured for basic domino with keeper, scheme proposed in [8] and the proposed circuit at different supply voltages. Table.2 and Chart 1 shows the PDP of the proposed circuit in comparison with the previous circuits. A plot of effect of supply voltage on delay is shown in Fig.6, and Fig.7 shows the simulated wave forms for a two input OR gate for various techniques. As can be seen from the wave forms the ripple in the Output is less in the proposed technique compared to the others thus reducing power dissipation. Output fall time is also less compared to other schemes, thus resulting in lower PDP

Table 1. Power and Delay comparison of the proposed scheme with existing schemes						
Supply voltage	Footless domino with keeper		Scheme of [8]		Proposed Scheme	
	Power	Delay	Power	Delay	Power	Delay
0.8v	2.55486E-06	1.4252E-09	7.49396E-07	2.15495E-09	1.40833E-06	1.26735E-09
0.9v	4.5318E-06	8.09909E-10	1.23397E-06	1.31192E-09	1.91414E-06	8.61012E-10
1.0v	6.79293E-06	6.00192E-10	1.74945E-06	1.02174E-09	2.40339E-06	7.18092E-10
1.1v	9.32268E-06	5.13905E-10	2.35983E-06	8.67323E-10	2.96814E-06	6.40169E-10
1.2v	1.18686E-05	4.34043E-10	3.16067E-06	4.44132E-10	3.48228E-06	2.10916E-10

Table 2. PDP comparison of the proposed scheme with existing schemes

Supply voltage	Footless domino with keeper	Scheme of [8]	Proposed Scheme
0.8v	3.64119E-15	1.61491E-15	1.78485E-15
0.9v	3.67035E-15	1.61887E-15	1.64809E-15
1.0v	4.07706E-15	1.78748E-15	1.72585E-15
1.1v	4.79097E-15	2.04673E-15	1.90011E-15
1.2v	5.15149E-15	1.40375E-15	7.34469E-16



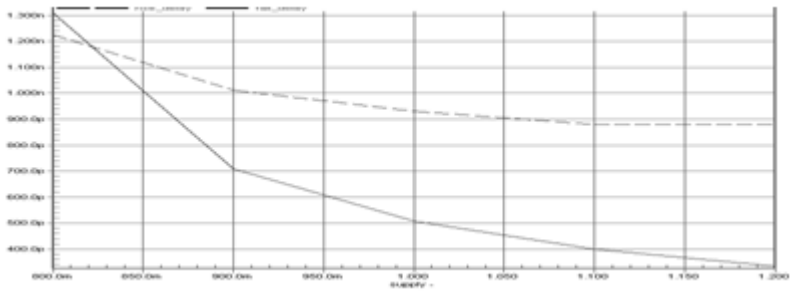


Fig. 6 Variation in Rise and Fall Delay for variation supply voltage

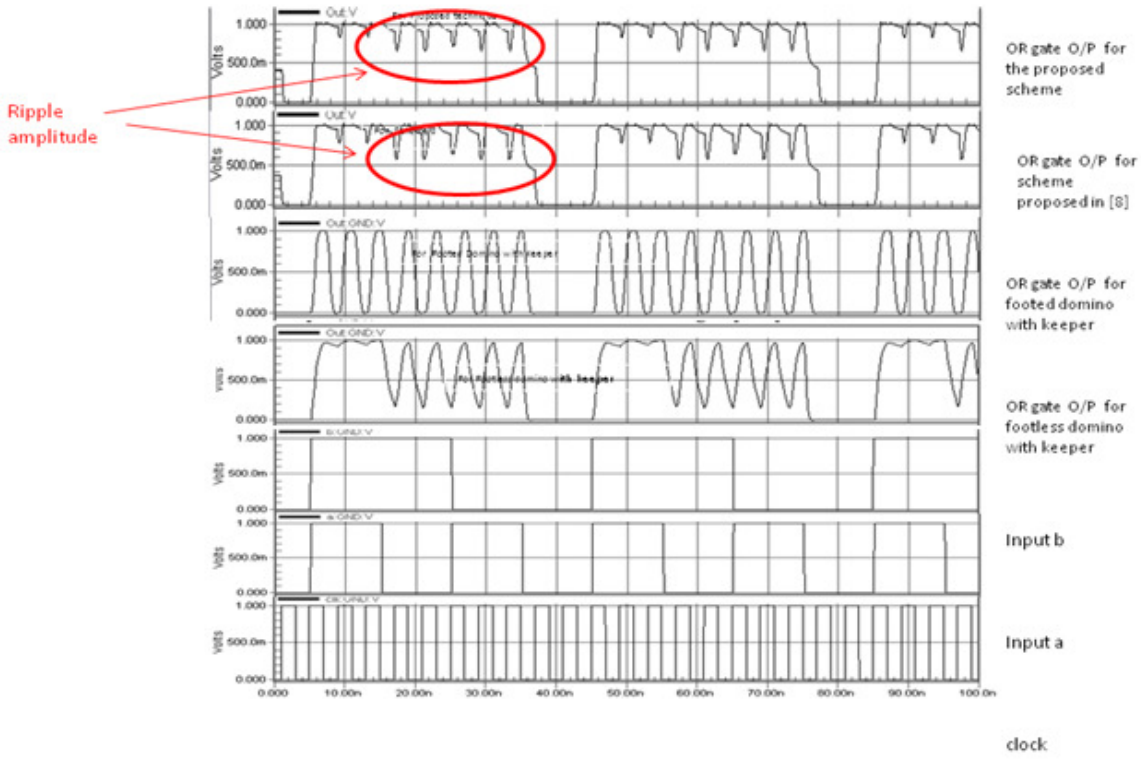


Fig.7 Simulation waveforms of Two input OR gate

5. CONCLUSION

In this paper we have proposed high speed low PDP domino logic circuit, which exhibits some noise tolerance at the output node. Simulations are done using Tanner T-Spice with PTM 16nm-low power technology files. From the results it is proved that the proposed design is better than the previous designs and offers about 29% reduction in delay and 3.5% reduction in PDP.

ACKNOWLEDMENTS

I would like to thank TEQIP-II for the facilities provided to carry out the work in the department.

REFERENCES

[1] John P. Uyemura “CMOS Logic circuit Design”Springer International Edition, 2005.
[2] H.L. Yeager et al, “Domino Circuit Topology”, U. S. Patent 6784695,Aug. 31, 2004.
[3] Krambeck, R.H., Lee.C.M., and Stephen Law.H.F., “High- speed compact circuits with CMOS”, IEEE J. Solid-State Circuits , 1982, 17, (3), pp 614-619.
[4] Oklobdzija, V.G., and Montoye, R.K., “Design performance tradeoffs in CMOSdomino logic “. Proc. IEEE Conf. on Custom Integrated Circiuts, May 1985, pp.334-337.
[5] Oklobdzija, V.G., and Montoye, R.K., “Design performance tradeoffs in CMOS domino logic “., IEEE J. Solid-State Circuits 1986, 21,(2), pp 304-306.
[6] Moahmoodi-Meimand H, Roy K., “ Diode-footed domino: a leakage tolerant high fan-in dynamic circuit design style” , IEEE Trans. Very Large Scalr Integr. Syst., 2004, 51, (3), pp. 495-503.

[7] Frustaci F.,Corsonellop., Cocorullo G., "A new noise tolerant dynamic logic circuit design", IEEE Ph.D. Research in Microelectronics and Electronics, PRIME 2007, Bordeaux, France, July 2007, pp 61-64.

[8] Preetisudha Meher, Kamala Kanta Mahapatra, "A technique to increase noise tolerance in dynamic digital circuits",Asia Pacific conference in post graduate Research in Microelectronics and Electronics(PRIMEASIA), December 2012, pp229-233

AUTHORS

Mrs. P. Koti Lakshmi is working as Assistant Professor in the department of ECE at Osmania University , Hyderabad. She has 15 years of teaching experience. She obtained her AMIETE Degree from IETE, New Delhi in 1999, M.E in Digital systems from Osmania University, Hyderabad in 2004, and currently pursuing Ph.D in VLSI Design from Osmania University , Hyderabad, Andhra Pradesh . Her areas of interest include VLSI Design and Wireless Communications.

Prof. Rameshwar Rao is Professor(Retd.) of department of ECE, University college of Engineering, Osmania University, Hyderabad. During his tenure he held many positions as Vice Chancellor of JNTUH, Hyderabad, Andhra Pradesh., Dean, Faculty of Engineering, Osmania University (OU)., Convener, PGECET. He obtained B.E. degree in ECE from OU, M.Tech. and Ph.D. degree from the prestigious IIT Bombay. His work experience spans across 35 years as R&D engineer at Avionics Design Bureau, Hindustan Aeronautics Ltd., Hyderabad and as an eminent teacher at Osmania University, Hyderabad. His research interests include VHDL Modeling & Synthesis, (During last three years), Data and Computer Communications, Detection and Estimation Theory, Information and Coding Theory, Microprocessor based applications and VLSI Design. He has to his credit more than 60 conference/ journal publications.

INTENTIONAL BLANK

AUTHENTICATION AND KEY AGREEMENT IN 3GPP NETWORKS

Krishna Prakash and Balachandra

Department of Information and Communication Technology
Manipal Institute of Technology, Manipal University, Manipal, India
kkp_prakash@yahoo.com
bala_muniyal@yahoo.com

ABSTRACT

The huge demand for mobile communications with broad band and usage of new wireless applications motivated the development of new wireless access technologies. The recent expansion of wireless technologies, novel applications and the advancement in mobile technology after UMTS-3G has been taken up to the next level by the 3GPP Long Tem Evolution/System Architecture Evolution (LTE/SAE). It has achieved the realisation of better bandwidth, full interworking with other access/backend systems using all-IP architecture with well-defined interworking with circuit switched system. The system is defined to work across multiple access networks (3GPP and non 3GPP) may be trusted or non-trusted. The security mechanism in wireless area has evolved from original analog systems through GSM and UMTS. The GSM has focussed the security for radio path whereas UMTS has enhanced it in to network functionalities. The future networks based on IP mechanism demands more security features, since the threats related to IP are also possible.

KEYWORDS

3GPP, LTE, SAE, UMTS, AKA

1. INTRODUCTION

Mobile computing provides flexibility of computing environment over physical mobility. The user of a mobile computing environment will be able to access the data, information or other logical objects from any device in any network while on the move. To make the mobile computing environment ubiquitous, it is necessary that the communication bearer is spread over both wired and wireless media.

The emerging mobile industry expected to be characterised by increasingly personalised and location based services. The availability of user preferred information despite of location made mobile computing successful. The advancement of mobile technology has revolutionised the way people use mobile devices in their day to day activity.

David C. Wyld et al. (Eds) : ACITY, DPPR, VLSI, WiMNET, AIAA, CNDC - 2015
pp. 143–154, 2015. © CS & IT-CSCP 2015 DOI : 10.5121/csit.2015.51313

Mobile computing offers a various services for the user over physical mobility. The user of a mobile computing environment will be able to access to data, information or other logical objects from any device in any network while on the move. To make the mobile computing environment ubiquitous, it is necessary that the communication bearer is spread over both wired and wireless media.

2. THE LTE/SAE 4TH GENERATION (4G) NETWORK SECURITIES ARCHITECTURE

Fig1 and Fig 2 demonstrate the LTE Network architecture. It is comprised of the Evolved Packet Core (EPC) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN). The EPC is an all IP and fully packet switched backbone network in LTE system. The IP Multimedia System takes care of voice service. When a User Equipment (UE) connects to the EPC, the Mobility Management entity takes care of mutual user authentication. It is equivalent to the Universal Terrestrial Radio Access Networks(UTRAN) Serving General Packet Radio Support Node (SGSN) enables the transfer of subscription and authentication data for the authentication and authorization of user access [1][2]. The Evolved-Universal Terrestrial Radio Access Network (E-UTRAN) consists of a node called eNode-B which has the functionality of node B and Radio Network controller of UTRAN and communicates with user equipment.

Following are some new functionality introduced by The LTE networks compared to the 3G wireless networks.

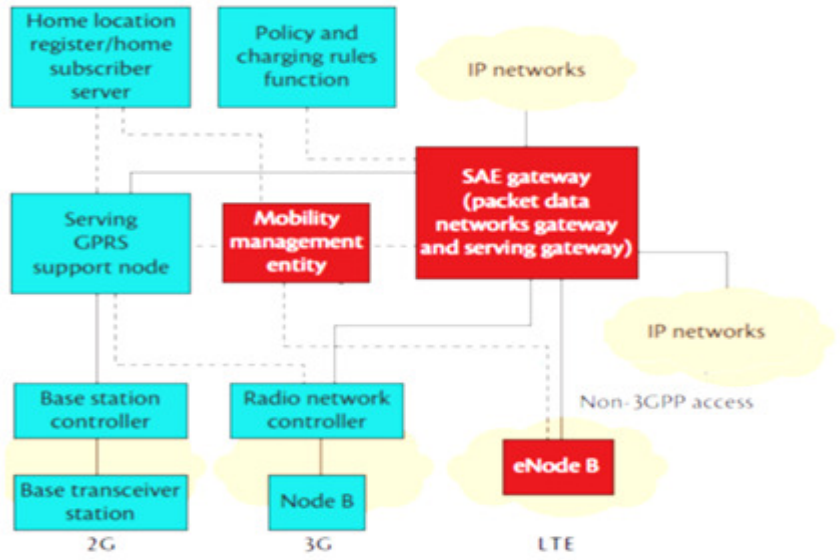


Fig 1: The SAE EPS architecture

1. A new type of base station called Home eNodeB (HeNB), and it is suggested by the 3GPP committee to improve the indoor coverage and capacity of the network. HeNB is a low power access point installed by the subscriber in the residence or small working areas to increase the coverage of voice and high speed data. It is connected to the EPC over internet via broad band backhaul [3].

2. In addition to the E-UTRAN, the LTE-A system supports non 3GPP access networks such as Wireless Local Area Networks(WLAN) and Code Division Multiple Access (CDMA) systems are allowed to connect to EPC. The two types of non 3GPP access networks namely trusted and non-trusted exists in use and for untrusted non 3GPP access networks the UE needs to pass a evolved packet data gateway(ePDG) connected to the EPC.

3. The LTE-A system supports a new type of data communication between entities called as Machine Type Communication (MTC) capable of data exchange without any human intervention. It is the communication between different devices (usually sensors) and the core network. The MTC user and the MTC Server are the two entities involve in the system and the MTC user uses the services provided by the one or more MTC servers for the operation of MTC devices. The MTC server is connected to the LTE network for the communication with Machine Type Communication Devices (MTCD)

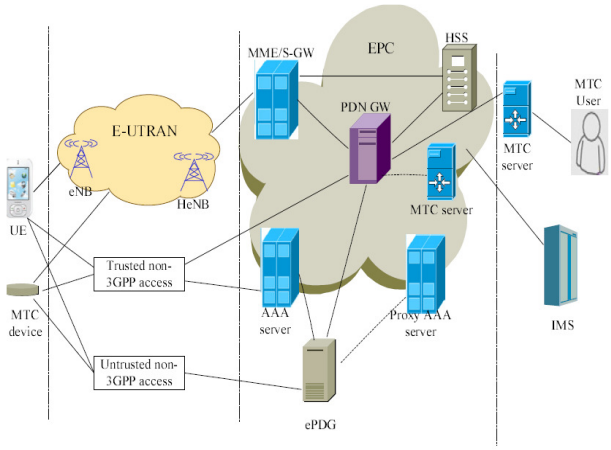


Fig 2: The LTE Network Architecture

3. LTE SECURITY ARCHITECTURE AND KEY HIERARCHY IN EPS

The basic security principles of smart phones and common PC are different. The device hosts multiple applications and allows the user to access internet irrespective of location. The complex software and infrastructure used in mobile device make the system more vulnerable and also the data exchanged between devices is a point of concern. The limited resources such as CPU and memory limit the sophistication of possible security solutions. A complex security algorithm that is used for real life applications cannot be directly ported and used in mobile devices.

3.1 LTE Security Architecture

The Fig 3 shows five different security levels defined by 3GPP committee for LTE architecture.

They are:

I .Network access security
II. Network domain security
III. User domain security
IV. Application domain security

V. Non 3GPP domain security

The following paragraphs briefly discuss the aforementioned functionalities.

I. Network Access Security

These security features facilitates the UEs for the secure access to EPC and protects possible attacks on radio link through integrity protection and ciphering between the USIM, ME, E-UTRAN and entities of EPC (both serving networks and home networks).

II. Network domain security

The set of security features protects possible attack on wire line networks and enables the data exchange in secure manner.

III. User domain security

The mutual authentication of USIM and ME is supported using a secret PIN before they can access each other.

IV. Application level security

These are the set of security features that enables the application in UE and the service provider domain for the secure exchange of messages.

V. Non 3GPP domain security

These are the set of features enables the UEs to securely access to the EPC via non 3GPP access networks and provide security protection on the access link.

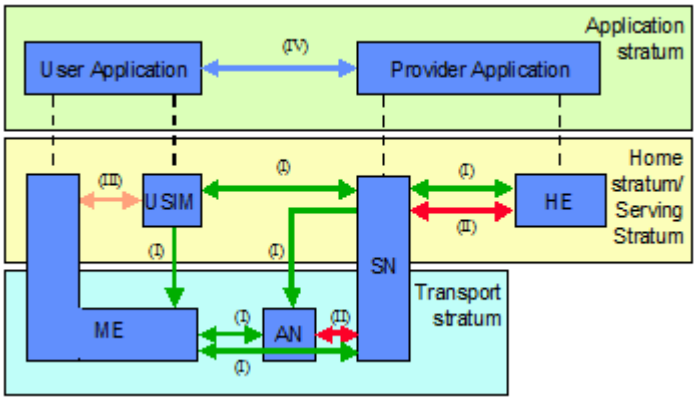


Fig 3: The LTE Security Architecture

3.2 Keys and Key Hierarchy

In the Evolved Packet Core Authentication and Key Agreement (EPS AKA) protocol, all the keys that are needed for various security mechanisms are derived from intermediate key K_{ASME} which is viewed as local master key for the subscriber in contrast to permanent master key K. In the

network side, the local master key K_{ASME} is stored in the MME and permanent master key is stored in the AuC [4]. This approach provides the following advantages.

1. It enables cryptographic key separation, where the usage of each key in one specific context and knowing one key does not deduce the second one.

2. The system is improved by providing key freshness and it is possible to renew the keys used in security mechanism. The EPS AKA is need not be run every time when the key to be renewed for protecting the radio interface and also the home network is not involved every time. This introduces a security versus complexity trade-off situation. For EPS, the security benefits of using an intermediate key overweigh the added complexity which was not true in 3G.

The base station eNB stores another key K_{eNB} and the addition of K_{eNB} makes it possible to renew keys for protection of radio access without involving MME.

3.3 Key Derivations

Figure 4 shows the hierarchy of keys used in EPS. The hierarchy contains one root key (K), several intermediate keys such as CK, IK etc. and a set of leaf keys [5]. The purpose of the different keys are explained below.

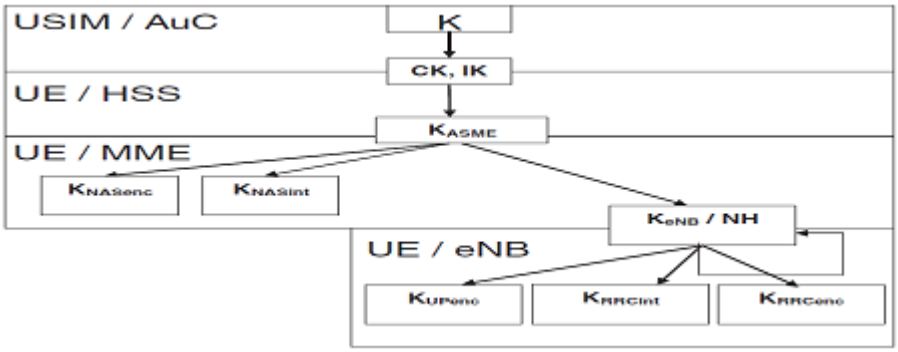


Fig 4: EPS Key Hierarchy

1. K is a random bit string and it is a subscriber specific master key stored in USIM and AuC.
2. CK and IK are 128 bit keys derived from K using additional input parameters.
3. K_{ASME} is derived from CK and IK using two additional parameters, the serving network id and bitwise sum of two additional parameters (SQN and AK from the EPS AKA procedure). The K_{ASME} serves as local master key.
4. K_{eNB} is derived from K_{ASME} and the additional input a counter. This additional parameter is needed to ensure that each new key K_{eNB} derived differs from the earlier key.
5. NH is another intermediate key derived from $K_{ASME,}$ and used in handover situations. It is derived from K_{eNB} for the initial NH derivation or previous NH as an additional input.
6. K_{RRCenc}, K_{RRCint} and K_{UPenc} are used for the encryption and integrity of RRC and Users.The complex key hierarchy achieves the key separation and prevents related key attack. The key hierarchy achieves key renewal very easily without affecting the other keys. When one key is changed, only the keys dependent on it have to be changed and others may remain same.

4. AUTHENTICATION AND KEY AGREEMENT PROTOCOL IN EPS (EPS AKA)

Authentication is a mechanism where the system verifies the identity of a user, who wishes to access it for availing some services. Mutual authentication is performed if both the communicating parties want to confirm each other. LTE/SAE architecture uses IP based mobility control technology and which has two components namely the access network and the core network. The access network is called evolved universal terrestrial radio access network (E-UTRAN) and the core network is called evolved packet core (EPC). Access security in E-UTRAN consists of following different components.

1. Mutual authentication between the network and UE.
2. Key derivation for ciphering and integrity protection.
3. Ciphering, integrity and replay protection of signalling between UE and MME, UE and enodeB.
4. Use of temporary identities in order to prevent the sending of permanent user identity over radio link.

4.1 User and Terminal Identification and Confidentiality

Similar kind of subscriber identity is used by 2G, 3G and EPS. It is composed of 3 parts (MCC, MNC, IMSI). Using IMSI the permanent authentication key K used in EPS AKA is identified. The IMEI is used to identify the terminals. The confidentiality of user identity against passive attacks is protected by assigning a temporary identity called TMSI in 2G and 3G. The EPS also adopts same mechanism to protect the actual UE id and it is called as Globally Unique Temporary UE Identity (GUTI). The GUTI has following components.

1. GUMMEI (Globally Unique MME Identifies) for the global unique identification of MME that allocated the GUTI.

2.M-TMSI uniquely identifies the UE within the MME that allocated the GUTI.

The GUMMEI is constructed from the MCC, MNC and MME identifier.

In GSM and 3G there is no user confidentiality protection from active attacks. In active attack, the attacker would use a device known as 'IMSI catcher', incorporates a false base station for sending an identity request message to UE [6]. The UE may respond with IMSI. The identity request is needed to recover from the cases where the network lost the association between temporary user identity and IMSI such as MME crash. Without such recovery the user could be permanently locked out from the system. 3GPP discussed the means by the use of public key certificates in UE. For roaming cases when the MME reside in another operators network may need the existence of public key infrastructure spanned across the operators with mutual agreement. In EPS, the UE not transmitting IMEI to the network upon network request before NAS security has been activated [7].

4.2 Authentication and Key Agreement (EPS AKA)

The EPS AKA mechanism is shown figure 5 and it is a combination of following three procedures.

1. A method to generate EPS Authentication vectors in the HSS up on request from the MME and distribute to MME.
2. A procedure to mutually authenticate and establish a new shared key between serving network and UE.
3. The mechanism to distribute authentication data inside and between serving networks.

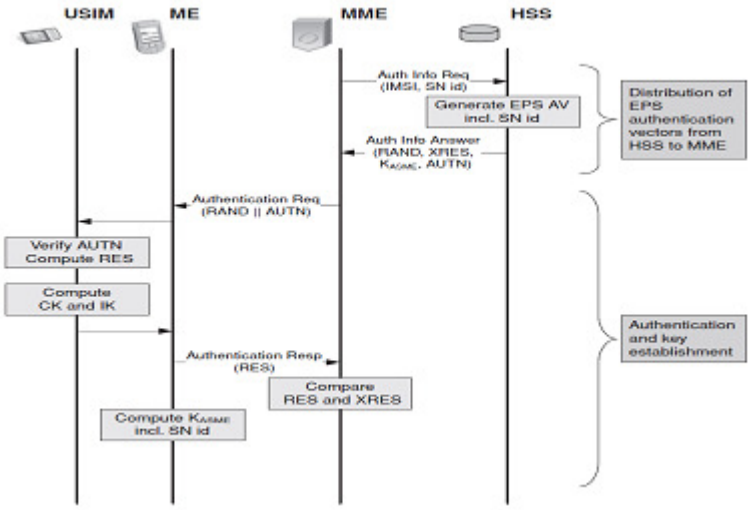


Fig 5: EPS AKA Protocol

4.3 Generation and distribution of Authentication Vector from HSS to MME

The MME invokes a procedure by requesting authentication vectors from the HSS. The authentication information request includes the IMSI, serving network id SN id of requesting MME with an indication of authentication information request from EPS. The SN id is required for the computation of K_{ASME} in HSS.

Up on the request of authentication vectors, the HSS may compute it or retrieve the pre computed values from the HSS database [8][9]. TS33.401 recommends the sending of one authentication vector at a time, because the need for frequently contacting HSS for fresh AV is reduced in EPS through the usage of local master key, K_{ASME} . Based on local master key and the keys derived from it an MME can offer secure services. The pre computed AVs are not usable when the user moves to a different serving network owing to the binding of local master key to the serving network id. Pre computation is useful when the next AV request is likely to be issued by an in the same serving network, when the user is in home network.

The Fig 6 demonstrates the authentication vector generation procedure. An EPS AV consists of a Random number RAND, an expected response XRES , local master key K_{ASME} and an authentication token AUTN. Both UMTS AV and EPS AV play a major role in EPS AKA. The HSS outside the AuC derives the K_{ASME} key from the cipher and integrity keys CK and IK. The AuC starts generating a fresh sequence number SQN and a random challenge RAND. For each user the HSS keeps track of counter $SQN_{HE.}$

An Authentication Management Field (AMF) is included in the authentication token of each authentication vector. The AuC computes following values after it receives a request from HSS.

1. Message Authentication Code (MAC) =f1(K||SQN||RAND||AMF), where f1 is a message authentication function.
2. Expected response, XRES = f2 (K||RAND), f2 is a truncated message authentication function.
3. A Cipher Key, CK = f3 (K||RAND), using a key generating function f3.
4. An Integrity Key, IK = f4 (K||RAND), f4 is a key generating function.
5. Anonymity Key, AK = f5 (K||RAND), f5 is a key generating function.
6. Authentication Token, AUTN = (SQN XOR AK) ||AMF||MAC.

If the operator decides no concealment of SQN, then f5=0 (AK=0).

After the receiving UMTS authentication vectors from AuC, using the Key Derivation Function, KDF and CK, IK, SNid, (SQN XOR AK) are used for producing K_{ASME}. The CK, IK are deleted from HSS and it is only used for the computation of EPS authentication vectors and not allowed to leave HSS.

The usage of AMF are

1. Indicating the algorithm and key used to generate a particular authentication vectors when multiple algorithms and permanent keys are used.
2. Change of parameter relating to SQN verification in the USIM.
3. Setting threshold values for key lifetimes.

The length of authentication parameters CK, IK, RAND and K all are 128 bit long and it is expandable up to 256 bits if needed in the future.

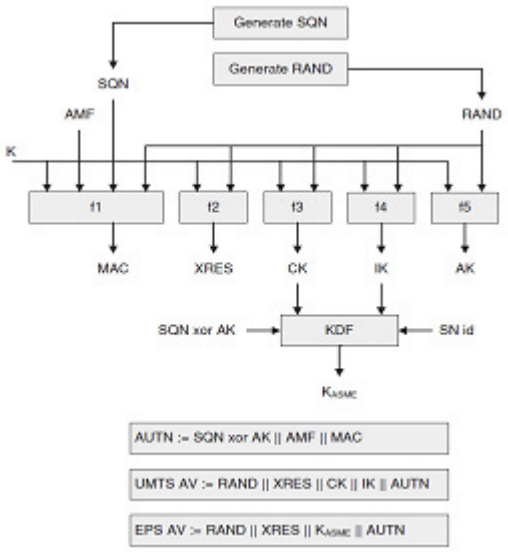


Fig 6: Generation of Authentication Vectors

4.4 Mutual Authentication and Establishment of a Shared Key between serving network and UE

After successful completion of the authentication of the user and generation of new local master key, K_{ASME} between MME and UE, the USIM performs verification of freshness of the authentication vector and authentication of its origin (User Home network). The K_{ASME} is used for the generation of further keys in subsequent procedures.

The MME invokes the procedure by using next unused EPS authentication vector in the MME database if more than one available. If the MME has no EPS AV, it requests one from the HSS. The MME sends a random challenge RAND and the authentication token for the network authentication AUTN from the selected EPS authentication vector to the mobile equipment which forwards it to the USIM. The USIM performs verification as follows.

The USIM first computes the anonymity key AK and retrieves the sequence number. The USIM next computes XMAC and verifies the MAC included in the AUTN as shown in Fig 7. The USIM verifies the retrieved sequences are in correct range or not for the satisfaction of following conditions.

1. Once the USIM has successfully verified an AUTN it shall not accept another AUTN with same sequence number to prevent the multiple usage of sequence number SQN.
2. It is required to allow out of order use of sequence numbers. Out of order usage of sequence numbers may occur when two different entities such as SGSN MSC/VLR request a batch of authentication vectors from HSS and use these in the interleaved fashion for AKA run with UE to reduce the synchronization failure rate.
3. The USIM may reject time based sequence numbers if it was generated too long ago.
4. The SQN verification mechanism may reject the SQN a jump from last successfully verified SQN is too big.

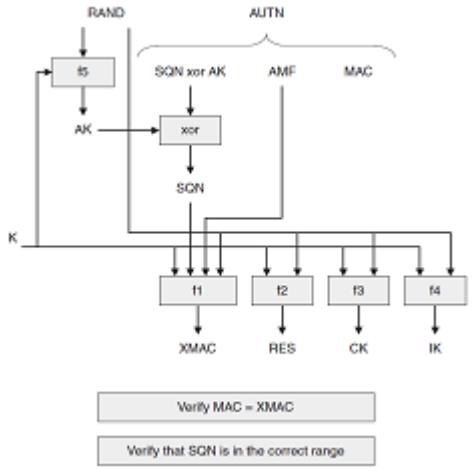


Fig 7: User Authentication Function in USIM

5. SECURITY THREATS AND VULNERABILITIES IN LTE

The EPS AKA scheme lacks a privacy protection under multiple instances of disclosure of IMSI[10]. When an UE registers to the network for first time/current MME cannot be contacted/ IMSI cannot be retrieved due to synchronization failure when it roams to new MME, the current MME or new MME requests the IMSI of UE as shown in messages 1 and 2 of Fig 8, then the UE must transmit IMSI in plaintext format and the disclosure of it may incur severe security problems [11][12]. Once the IMSI is captured, the adversary may attempt to tamper the information, subscriber or location information and then disguise the real UE and may launch attacks such as DOS to destroy the network.

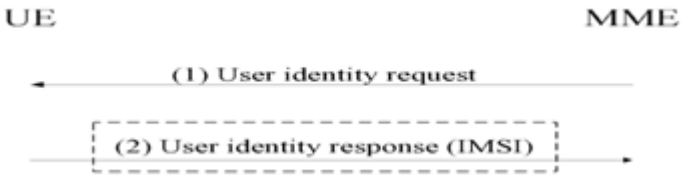


Fig 8: IMSI request process

The EPS AKA scheme cannot prevent DoS attacks. The MME forwards UEs request to the HSS/AuC before the UE has been authenticated by the MME as shown in message 3, and MME can only authenticate the UE after it receives RES from UE as shown in Fig 9. Using these two conditions the opponent can launch DoS attack between HSS and MME. The attacker can disguise a legitimate UE to constantly send fake IMSI to overwhelm HSS/AuC. Because of this, HSS consumes its computational power for the generation of authentication vectors for UE and MME consumes its memory buffer to wait overly long period of time for a legitimate or false response from the corresponding UE.

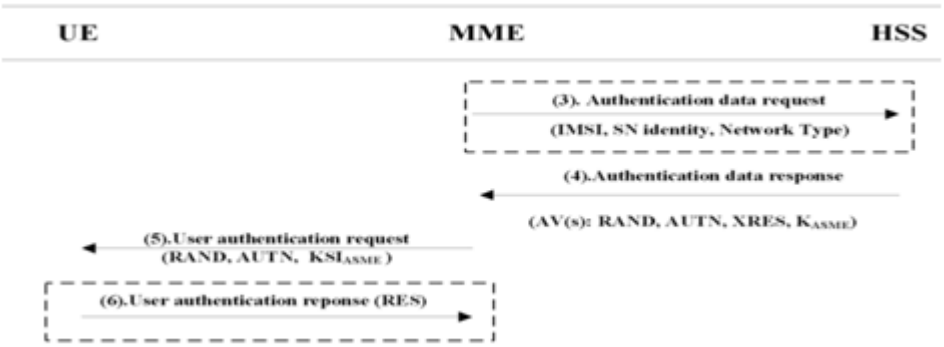


Fig 9: Authentication Data request and Mutual Authentication

In the EPS AKA as shown in Fig 9, the SN turns to HN for request of another set of authentication vectors when the UE stays in the SN for longer period of time and exhausts its set of authentication vector for authentication results in the bandwidth consumption and authentication signalling overhead between SN and HN, storage consumption in the SN.

The EPS AKA protocol is a delegated protocol, where almost all authentication authorities are delegated from home network to visited network requires strong trust relationship between

operators. The increased number of roaming entities and other access systems, in the heterogeneous networks the trust assumptions seems to be outdated [13].

5.1 Other Vulnerabilities in LTE/SAE

The vulnerabilities in LTE/SAE is classified under the following categories [14].

1. Threats against user identity and privacy
2. Threats of USIM/UE tracking
3. Threat related to handovers and base stations
4. Threats related to denial of service
5. Threats of unauthorised access to the network
6. Compromise of eNB credentials and physical attack on eNB
7. Attacks on core networks, including eNB location based attack

6. CONCLUSION

The last few years have witnessed a huge growth in wireless mobile industry. In the near future it can be expected that the mobile terminals are treated like internet terminals. The number of users using smart phones for various purposes has been increasing tremendously for m-commerce applications.

The mobile network security needs to be adopted from the entire mobile network point of view rather than a single device. From the history of mobile communication, attempts have been made to reduce a number of technologies to a single global standard. The first two generations has fulfilled basic mobile voice with capacity and coverage. Third generation opened the gate for the experience of high data transfer speed followed by fourth generation with added functionalities. The fifth generation is under research, aiming for high quality of service with more speed.

REFERENCES

[1] Gunter Schafer, “Security in Fixed and Wireless Networks-An Introduction to Securing Data Communication”, Second Edition, Wiley Publishers.
[2] Randall K Nichols et al., “Wireless Security-Models, Threats, and Solutions”, TaTa Mcgraw Hill Edition.
[3] 3rd Generation Partnership Project- Technical Specification Group Services and System aspects: Service requirements for Home Node (HNB) and Home eNode B(HeNB) (Rel 11), 3GPP TS 22.220 V11.6.0 Sep. 2012.
[4] Masoumeh Purkhiabani et al., “Enhanced Authentication and Key agreement Procedure of Next Generation Evolved Mobile Networks”, IEEE Journal, 2011.
[5] Jin Cao et al., “A Survey on Security Aspects of LTE and LTE-A networks”,IEEE Communications Survey and Tutorial, Vol. 16, No 1, Frist Quarter 2014.
[6] Sankaran C.B. “ Network Access Security in Next Generation 3GPP Systems”, IEEE Communications Magazine, February 2009.
[7] Ivan Stojmenvovic ,“ Handbook of Wireless Networks and Mobile Computing” , ISBN:0-471-41902-8.
[8] Hakima Chauchi ,“ Wireless and Mobile Network Security”, John Wiley Publications.
[9] Xinghua Li et.al., “A USIM based uniform access authentication framework in mobile Communications” Research Article,EURASIP Journal onWireless Communications and Networking, 2011.

[10] Li Xiehua "Security Enhanced Authentication and Key Agreement protocol for LTE/SAE Network" IEEE Journal, 2011.
[11] Zahra Ahmadian et al., "Security Enhancements against UMTS-GSM interworking attacks" ScienceDirect, 2010.
[12] Mariantonietta La Polla et al., " A Survey on Security for Mobile devices" IEEE Communications Surveys and Tutorials", 2013.
[13] Anastasios N. Bikos "LTE/SAE Security issues on 4G Wireless Networks", IEEE Computer and reliable societies, 2013.
[14] Inhyok cha et al., "Trust in M2M Communications-Addressing New Security Threats" IEEE Vehicular Technology Magazine, 2009.

MOBILE-BASED VIDEO CACHING ARCHITECTURE BASED ON BILLBOARD MANAGER

Rajesh Bose[1], Sandip Roy[2] and Debabrata Sarddar[3]

[1, 2, 3]Department of Computer Science & Engineering,
University of Kalyani, Kalyani, West Bengal, India
[1]bose.raj00028@gmail.com, [2]sandiproy86@gmail.com,
[3]dsarddar1@gmail.com

ABSTRACT

Video streaming services are very popular today. Increasingly, users can now access multimedia applications and video playback wirelessly on their mobile devices. However, a significant challenge remains in ensuring smooth and uninterrupted transmission of almost any size of video file over a 3G network, and as quickly as possible in order to optimize bandwidth consumption. In this paper, we propose to position our Billboard Manager to provide an optimal transmission rate to enable smooth video playback to a mobile device user connected to a 3G network. Our work focuses on serving user requests by mobile operators from cached resource managed by Billboard Manager, and transmitting the video files from this pool. The aim is to reduce the load placed on bandwidth resources of a mobile operator by routing away as much user requests away from the internet for having to search a video and, subsequently, if located, have it transferred back to the user.

KEYWORDS

Video-on-Demand, Video Streaming, Multimedia Cache, Data Center, 3G Network

1. INTRODUCTION

Video streaming services are extremely popular today. Today, these services are no longer confined to wired network but are also delivered wirelessly. A growing number of content providers, e.g., YouTube, Netflix, etc. are targeting mobile device users whose numbers are only expected to grow by leaps and bounds. While caching is commonly used by wireless service providers to improve the streaming quality, our proposed Billboard Manager model has been designed keeping in mind cellular and mobile networks with the aim of delivering video content more efficiently over available bandwidth. The proposed Billboard Manager works on focusing on techniques of converting videos to appropriate codecs and resolutions fit for a given mobile device from which request for the video originates. With the Billboard Manager installed, the requested video would be located from a cloud-based node nearest to the location of the user and then stored in its cache. Depending on the model of the mobile device, the Billboard Manager

David C. Wyld et al. (Eds) : ACITY, DPPR, VLSI, WiMNET, AIAA, CNDC - 2015
pp. 155–162, 2015. © CS & IT-CSCP 2015 DOI : 10.5121/csit.2015.51314

would then transcode the video and begins transmitting it to the mobile device taking into consideration the current connectivity quality [1]. Caching service has been around for some time now. Various techniques exist of cache implementation. By itself, caching is not a complex concept to understand. A cache stores information and data which is frequently accessed and provides fast transfer rates by focusing on optimizing transmission rates. Caches are extensively used on every conceivable electronic gadget or device today. The importance of caching cannot be over-emphasized going by the several hundred million personal computing devices which use processor-based caches [2].

1.1. How caching service of Billboard Manager works

We propose to position The Billboard Manager between mobile device clients and cloud video servers. As soon as a request for video is placed to the mobile content service provider, the Billboard Manager scans its cache for the video requested. If the video exists in its cache, the Billboard Manager begins transcoding of the video in the appropriate format suitable for playback on the mobile device requesting the video. Following completion of the process, the Billboard Manager would then begin transferring the video to the mobile device. In case the video does not exist in its cache, the Billboard Manager attempts to locate it from its nearest cloud servers hosting videos. In case the video is found, the Billboard Manager caches it in dedicated cache repository before transcoding and transmitting the video to the mobile device user. As a result of delivering cached content, the mobile content service provider can reduce the quantum of network bandwidth consumption on the server-side, while enhancing end-user experience by offering superior response times to users' requests. Our proposed Billboard Manager has been designed to deliver the following benefits through video caching in the following manner [3]:

1.2. Quality of experience is optimized:

Subscribers are able to enjoy the best possible quality of video that are processed for prevailing mobile network conditions.

1.3. Bandwidth costs are lowered:

Caching of videos that are looked up by users more than once, translates to reduction in the number of times the videos are looked up and transferred from the respective sites providing such content. This, in turn, decreases the volume of data traversing the networks.

1.4. Enhanced network efficiency:

The effect of caching videos significantly improves network conditions by lowering the volume of over-the-top videos transmitted even during peak hours. Consequently, network traffic and/or services that are non-cacheable also see an upward trend in performance.

1.5. Start up to the playback times is faster:

Video start times are improved as the content is moved closer to the subscriber. Wait times for the content to be transmitted from distant nodes are also cut down.

1.6. Video stall reduction:

Delays arising out of transmission from content servers have little impact on user experience. Video caching services ensure that requested videos are transferred to the respective users over the shortest distance.

1.7. Streaming Video over 3G Mobile Networks through Billboard Manager

Our proposed Billboard Manager would employ reduction techniques of reducing resolution size from CIF to QCIF and using fewer frames per second. However, the efficiency of our proposed model would rest not on video compression techniques alone, but on dynamically sending blocks of video in contiguous blocks. The method has been demonstrated by [4].

2. RELATED WORK

Video caching is a new research area that has not been sufficiently explored. During recent years, a few commercial video caching systems have been developed. Another body of related work is in the area of scalable video-on-demand systems [5-8]. The idea is to reduce server load by grouping multiple requests within a time-interval and serving the entire group in single stream .The Middle Man architecture is a collection of cooperative proxy servers that collectively act as a video cache for a well-provisioned local area network [9]. Video streams are stored across multiple proxies where they can be replaced at a granularity of a block .They examine performance of the Middle Man architecture with different replacement policies. Other related work has been done on memory caching for multimedia servers [10, 11]. While the basic principle of caching data in different memory levels of a video server has some similarities with storing data in a distributed caching system, there is a fundamental difference. The spatial distance between different memory levels in a server is zero. In contrast, spatial distance between distributed caching systems is not negligible and, therefore, has to be considered in the design of web cache management policies. A good report has been presented in [12]. Another good report on mobile cloud architecture that helps us to solve the caching problem is presented in [13]. Ref. [14] a proxy caching mechanism is used for improving delivered quality of layered encoded multimedia streams.

3. PROPOSED WORK

A mobile user sends request for video to the Billboard Manager. The request is checked against its index table. If the request for content is not found in its index table, the Billboard Manager passes an HTTP byte range request to retrieve the content from one of the registered cloud nodes of the Billboard Manager. While retrieval, a copy of this video is saved in its own database to serve identical requests in future. Alternatively, if the request for video can be served directly, the Billboard Manager is able to deliver the video direct from its own cached resource thereby reducing the processing time which would have been otherwise expended in searching. The Billboard Manager also uses a video converter which is a part of its own system. The transcoder transforms video from online video streaming sites, e.g., YouTube, Daily motion, etc., into the requisite format playable on mobile devices. The Billboard Manager also splits the encoded streams into segments and sends each converted stream segments to mobile devices over mobile or local networks. The Billboard Manager acts as an interface between users of mobile devices and the cloud. Our proposed model involving Billboard Manager has been designed keeping in mind saving of 3G network bandwidth at either the user and/or the cloud service provider ends.

The sole purpose of our proposed model is to achieve a considerable reduction in data transmission across any given 3G mobile networks from cloud-based nodes. The end goal is to achieve direct savings in terms of costs and bandwidth resources utilized. Our model suggests an architecture, wherein, the Billboard Manager is able to provide video streams directly to a 3G mobile device user from its own cached resource rather than accessing the same video from across one of its own registered cloud nodes.

This directly translates in time savings as well. In the absence of a video not in its cached resource, the 3G mobile user requesting the service would be served with a controlled video stream. The mechanism of this would be managed in a manner such that the Billboard Manager initially begins caching chunks of the video from its own cloud-based nodes, and then sends from it, smaller streams to the 3G mobile device user. In this manner, the 3G mobile device user, after experiencing an initial time delay from the moment the user places the request to the time when the video starts on the device, enjoys an uninterrupted video stream thereon. Further, once a video is fully cached, the Billboard Manager can serve it from its resource to other mobile device users placing a request for it. This reduces overall bandwidth consumption as the Billboard Manager does not have to engage resources to obtain the video once again from its registered cloud-based nodes. The 3G mobile device user requesting the video only has to use their own (3G) network bandwidth and can do so with maximum effect with little or no delay between placing the request to the time the video starts playing. Thus, with the aid of this proposed model, a considerable amount of resources in terms of bandwidth, cost and time can be saved. Figure 1 shows the function of our proposed model.

Figure 1. Proposed Caching Architecture

3.1. Algorithm of our proposed cloud architecture

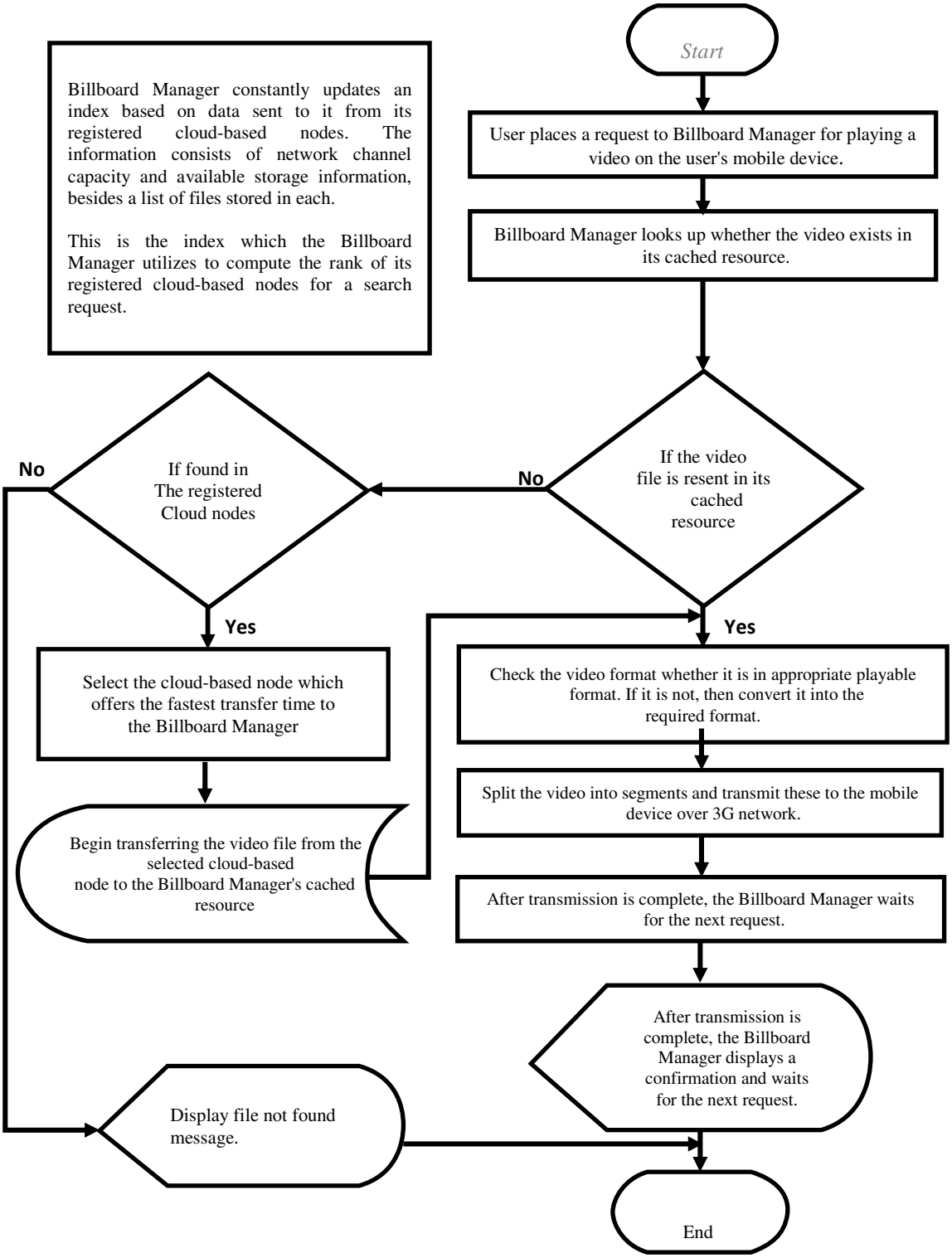
1. User places a request to play a video.
2. The Billboard Manager takes in the request. All such requests, in our proposed model, is routed through the Billboard Manager.

3. The request is processed against an index maintained by the Billboard Manager. The index contains details of videos that can be served by the Billboard Manager from its own cached resources and those from its registered cloud-based nodes.
4. If the requested video is found in its own cached resources, the Billboard Manager begins streaming it to the user requesting the video after processing it as described in the following steps #5 through 8. If not, the Billboard Manager moves forward to step #11 and #12.
5. The Billboard Manager checks whether the video is in the appropriate format required for playing on the mobile device of the user requesting the video.
6. If it is, it proceeds to step #8. If not, it proceeds with step #7, i.e., the following step.
7. The Billboard Manager converts the video to an appropriate format that can be streamed directly to the mobile device for instant playback.
8. The Billboard Manager splits the encoded video into segments. Each such segment is then transmitted to the mobile device over 3G mobile network.
9. After complete transfer of the file, the Billboard Manager stands by for the next request from a mobile device user.
10. If the requested video is not listed in its index, the Billboard Manager begins to lookup another set of index which details availability of the video at the registered cloud-based nodes of the Billboard Manager.
11. If the requested video cannot be found either in its cached resources or with any of its registered cloud-based nodes, the Billboard Manager notifies the user, skips all the following steps and awaits further request from the mobile device user. Otherwise, it proceeds with the next step of retrieving the video file from its registered cloud-based nodes.
12. If the Billboard Manager finds the video to be stored at more than one cloud-based node, it begins to select the best among them from where the video file can be transmitted to the Billboard Manager. The process of selection is based on several factors which are outlined in the next steps.
13. All the cloud-nodes registered with the Billboard Manager keep sending information at periodic intervals. The information consists of network channel capacity and available storage space.
14. The Billboard Manager constantly computes an index score for each of the registered cloud-based nodes. The index score is computed taking into account the constant periodic inputs from the nodes themselves, and the shortest route to each of the nodes.
15. At a given time when the Billboard Manager is about to select the best fit cloud-based node from where the video file is to be streamed back to it, the following steps are followed which are inherent to the steps #13 and 14 above. In other words, an instant index score is calculated based on the following parameters and comparisons.

    a. Select the shortest route to the cloud-based nodes hosting the video file.
    b. If more than one node shares identical route times, the channel capacity is compared. Otherwise, transmission of the video file begins from the cloud-based node to the Billboard Manager.
    c. After channel capacities are compared, the one with the maximum value is selected for the video file to be transmitted. Otherwise signal strengths are compared.
    d. The node with the best signal strength is selected and the video transmission takes place between the corresponding cloud-based node and the Billboard Manager.
    e. In case, after going through all the comparisons, more than one cloud-based nodes are identically matched to transmit the video file, one of them is chosen at random by the Billboard Manager.

16. Upon completion of transfer of the video, the Billboard Manager proceeds with step #5 onward.
17. The algorithm ends with the Billboard Manager completing the workflow cycle at steps #9. In the event, the video file does not exist within its registered network of cloud-based nodes, or within its own cached resources, step #11 is executed.

3.2. Flowchart of our proposed cloud architecture



4. CONCLUSIONS

This paper describes a proposed design of mobile video architecture based on our projected Billboard Manager. The algorithm detailed in this paper explains the steps that the Billboard Manager intends to take to create an efficient environment within which both the service provider and the user are able to optimize available bandwidth resources to the extent possible. Our proposed model would help send video streams at appropriate frame rates in a relatively short span of time, suitable for the mobile device requesting the video file. This is achieved with the help of caching and encoding services. In the initial stage, when the requested video is not available in the local cache, there would be a certain amount of time which would be involved to fetch it from the appropriate cloud node registered with our proposed Billboard Manager. Once the video is fetched, a copy of it would be stored automatically in its local cache by the Billboard Manager to serve subsequent requests. As a result, many resources in the form of bandwidth and time can be saved in cases where requests for the identical video file are ever placed in future. This proposed model wherein the Billboard Manager utilizes local cache database instead of repeatedly engaging bandwidth necessary to transfer files from cloud nodes to mobile device users, would help efficiently reduce video playback lag from the time it takes for the user to complete placing the request for the video.

ACKNOWLEDGEMENTS

The authors express their gratitude towards staff and members of the Department of Computer Science & Engineering, University of Kalyani for helping in arranging computation resources that have been used in the work.

REFERENCES

[1] Wu, Y., Zhang, Z., Wu, C., Li, Z. & Lau, F. (2013) "CloudMoV: Cloud-based Mobile Social TV", IEEE Transactions on Multimedia, Vol. 15, No. 4, pp. 821-832.

[2] Lanjewar, R., Sambare A.S. & Jain S. R. (2014) "A Survey on Peer to Peer sharing using Cloud Based Mobile Social TV (Cloud MoV)", International Journal of Research in Computer and Communication Technology, Vol. 3, No. 2, pp. 236-240.

[3] "ByteMobile Video Caching," –citrix.com/bytemobile Application Brief https://www.citrix.com/content/dam/citrix/en_us/documents/products-solutions/bytemobile-video-caching.pdf?accessmode=direct

[4] Dan, A., Sitaram, D. & Shahabuddin, P. (1996) "Dynamic batching policies for an on-demand video server", Multimedia Systems, Vol. 4, No. 3, pp. 51–58.

[5] Hua, K. A., Cai, Y. & Sheu, S. (1998) "Patching: a multicast technique for true video-on-demand services", in Proceedings of ACM Multimedia '98, Bristol, England, pp. 191-200.

[6] Aggarwal, C. C., Wolf, J. L. & Yu, P. S. (1996) "On optimal batching policies for video-on-demand storage servers", in Proc. of International Conference on Multimedia Systems'96, pp. 253-258.

[7] Sheu, S., Hua, K. A. & Tavanapong, W. (1997) "Chaining: a generalized batching technique for video-on-demand systems", in Proceedings of IEEE International Conference on Multimedia Computing and Systems, Ottawa, Ontario, Canada, pp. 110-117.

[8] Acharya, S. & Smith, B. C. (2000) "Middleman: A video caching proxy server", in Workshop on Network and Operating system support for Digital Audio and Video, June 2000.

[9] Dan, A. & Sitaram, D. (1996) "A generalized interval caching policy for mixed interactive and long video environments", in Proceedings of IS&T SPIE Multimedia Computing and Networking Conference, a Jose, CA, pp. 699-706.

[10] Dan, A. & Sitaram, D. (1997) "Multimedia Caching Strategies for Heterogeneous Application and Server Environments", Multimedia Tools and Applications, Vol. 4, No. 3, pp. 279-312.
[11] Sen, S., Rexford, J. & Towsley, D. (1999) "Proxy prefix caching for multimedia streams", in Proceedings of IEEE Infocom'99, New York, USA, pp. 1310-1319.
[13] Sarddar, D. & Bose, R. (2014) "A Mobile Cloud Computing Architecture with Easy Resource Sharing", International Journal of Current Engineering and Technology, Vol.4, No.3, pp. 1249-1254.
[14] Rejaie, R., Handley, M., Yuand, H. & Estrin, D. (1999) "Proxy Caching Mechanisms for Multimedia Playback Streams in the Internet", in Proceedings of the 4th International Web Caching Workshop, San Diego, CA, pp. 1-10.

AUTHORS

Rajesh Bose is currently pursuing Ph.D from University of Kalyani. He is an IT professional employed as Senior Project Engineer with Simplex Infrastructures Limited, Data Center, Kolkata. He received his degree in M.Tech. in Mobile Communication and Networking from WBUT in 2007. He received his degree in B.E. in Computer Science and Engineering from BPUT in 2004. He has also several global certifications under his belt. These are CCNA, CCNP-BCRAN, and CCA (Citrix Certified Administrator for Citrix Access Gateway 9 Enterprise Edition), CCA (Citrix Certified Administrator for Citrix Xen App 5 for Windows Server 2008). His research interests include cloud computing, wireless communication and networking.

Sandip Roy is currently pursuing Ph.D from University of Kalyani. He is an Assistant Professor in the Department of Information Technology, Brainware Group of Institutions, Kolkata, West Bengal, India. He has completed M.Tech in Computer Science & Engineering from HIT under WBUT in 2011. He has also done his B.Tech in Information Technology from WBUT in 2008. His main areas of research interest are Cloud Computing, Data Structure and Algorithm.

ebabrata Sarddar is an Assistant Professor in the Department of Computer Science and Engineering, University of Kalyani, Kalyani, Nadia, West Bengal, INDIA. He has done PhD at Jadavpur University. He completed his M. Tech in Computer Science & Engineering from DAVV, Indore in 2006, and his B.E in Computer Science & Engineering from NIT, Durgapur in 2001. He has published more than 75 research papers in different journals and conferences. His research interest includes wireless and mobile system and WSN, Cloud computing.

DROIDSWAN: DETECTING MALICIOUS ANDROID APPLICATIONS BASED ON STATIC FEATURE ANALYSIS

Babu Rajesh V, Phaninder Reddy, Himanshu P and Mahesh U Patil

Centre for Development of Advanced Computing
cdac.in

ABSTRACT

Android being a widely used mobile platform has witnessed an increase in the number of malicious samples on its market place. The availability of multiple sources for downloading applications has also contributed to users falling prey to malicious applications. Classification of an Android application as malicious or benign remains a challenge as malicious applications maneuver to pose themselves as benign. This paper presents an approach which extracts various features from Android Application Package file (APK) using static analysis and subsequently classifies using machine learning techniques. The contribution of this work includes deriving, extracting and analyzing crucial features of Android applications that aid in efficient classification. The analysis is carried out using various machine learning algorithms with both weighted and non-weighted approaches. It was observed that weighted approach depicts higher detection rates using fewer features. Random Forest algorithm exhibited high detection rate and shows the least false positive rate.

KEYWORDS

Mobile Security, Malware, Static Analysis, Machine Learning, Android

1. INTRODUCTION

Android is a widely used mobile platform and due to it's dominance in consumer space, Android becomes a lucrative target for malware developers who are exploiting the popularity and openness of Android platform for various benefits. Malware developers use Android marketplaces as entry points for hosting thesir malicious applications into the android user space. According to RiskIQ [1] report, malicious applications in Play store have grown by 388 percent from 2011 to 2013, while the number of such applications removed annually by Google has dropped from 60 percent in 2011 to 23 percent in 2013. As a large number of applications are uploaded and updated regularly on these market places, Manual analysis of all the applications is difficult task. Scarcity of effective mechanisms to detect these malicious samples has fueled the rise of malware applications on Android market places. In this regard we present DroidSwan, a system for classifying applications as malware or benign, based on static analysis of Android APK. DroidSwan extracts various crucial features from an Android application, assigns weight to these features and builds a classifier model using machine learning algorithms. The classifier model is trained using the malware data set of 1260 malware acquired from Genome Malware

David C. Wyld et al. (Eds) : ACITY, DPPR, VLSI, WiMNET, AIAA, CNDC - 2015
pp. 163–178, 2015. © CS & IT-CSCP 2015 DOI : 10.5121/csit.2015.51315

Project [2] and popular benign applications obtained from Google Play Store. The model was then tested against 500 malware samples obtained from Virustotal malware intelligence service [3].

Android applications are installed on to a device using an Android application package (APK) file. In our analysis, This APK file is disassembled for extraction of necessary features which form feature set to be used during classification. Figure 1 depicts how features are extracted from an APK. The application resources such as XML layout-definition files and images are stored in the 'res' directory which the malware writers use to inject malicious binaries like '.sh', '.elf' or '.exe' inside the images or other resource files used by the application. In most of the cases, these malicious binaries are found embedded within image files (.jpg and .png) used by the application. The AndroidManifest.xml file contains the name, version number and access rights of the APK. AndroidManifest.xml is a binary XML file. ApkParser [4] can be used to convert this binary XML file into a readable XML file. A malware application may hide some of the permission it uses by not declaring them in the Manifest file.

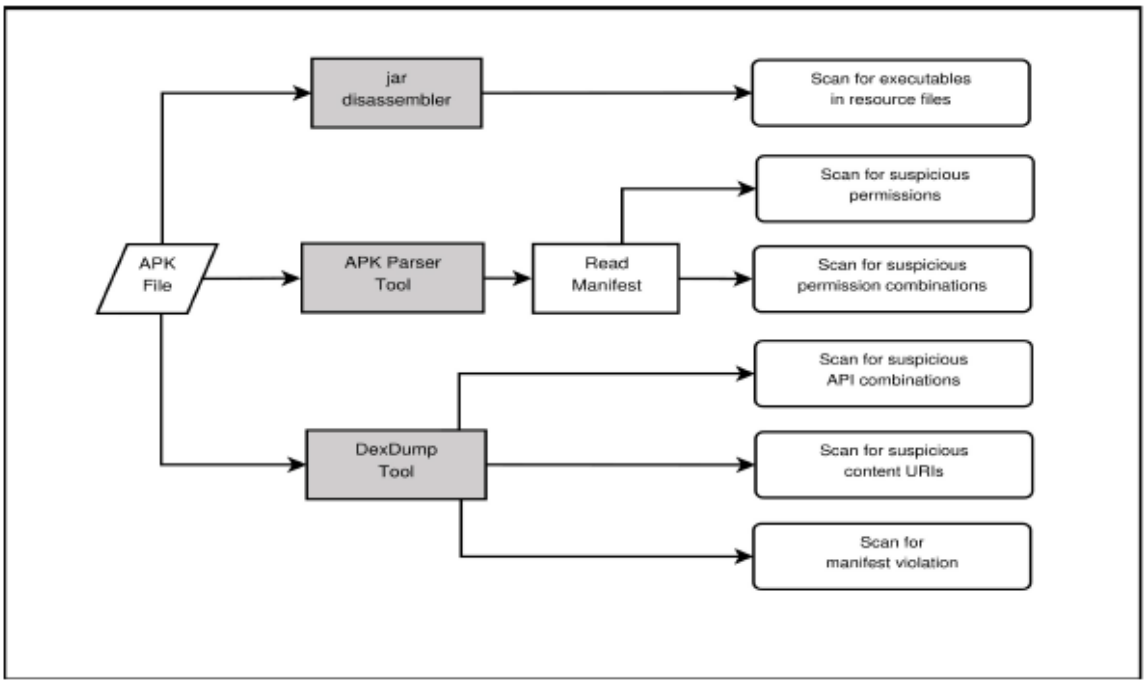


Figure 1. Feature extraction in DroidSwan

2. RELATED WORK

Androguard [5] statically extracts features from APK, but this tool shows high false positive rate as 80 out of 100 popular benign samples analyzed were assigned high androrisk score. Aubrey-Derrick Schmidt et al. [6] extracted function calls of an installed application using readelf command. These function calls were later compared with function calls of the malware executables present on a Remote Detection Server. In contrast to this, our approach does not analyze applications on an Android device because of limited resources like power, memory and data usage, but if needed it can be ported onto a mobile device. DroidRanger [7] detects

malicious applications of known malware families in popular Android marketplaces using permission-based behavioral foot printing. To detect malware from unknown families, DroidRanger uses heuristic-based filtering scheme. The drawback of DroidRanger is the requirement of manual operations while analyzing and collecting behavior of applications.

DroidMat [8] combines static and dynamic analysis approaches. It extracts features like permissions and intents using static analysis and API calls using dynamic analysis. In contrast to this, we perform static analysis to extract all the necessary features as a tool based on static analysis can be deployed on a gateway device with greater ease as compared to tool based on dynamic analysis. Adrieene et al. [9] proposed an approach to identify over privileged applications by comparing API calls invoked with permissions declared in the Manifest. William Enck et al. [10] proposed an approach where a certificate is generated during an application's installation. This certificate gives complete information about the application by rating them using Kirin security rules which are based on the combinations of permissions extracted from Manifest file. DroidAnalytics [11] is a signature based system for detecting repackaged applications. The drawback of this technique is it requires large and balanced data set of malware and benign samples. Shabtai et al. [12] applied machine learning classifier techniques like decision tree, Naive Bayes (NB), Bayesian Networks (BN) etc. to classify Android applications as games and utilities citing the non availability of malware applications.

They collected around 22,000 features initially and later reduced to 50 features for the purpose of classification. Our approach uses 24 features for classification.

3. APPROACH

This section explains our approach. The following subsections describe feature selection for feature set, weight assignment to the features, selection of feature vector and finally the working of DroidSwan.

3.1. Features

3.1.1. Suspicious Permissions and Permission Combinations

A permission is a restriction limiting the access of an application to the device to protect critical data and code that could be misused to distort or damage the user experience. We considered the patterns of suspicious permissions in malware samples as discovered by Y.Zhou et.al. [13]. For extracting permissions used by an application we use APKParser tool. The permissions extracted were analyzed and cross verified for high occurrence across malware samples available in our training dataset. Out of all the permissions specified as suspicious by Y.Zhou et.al, we discarded those permissions which were present in large numbers in benign samples as these would not significantly contribute during classification process. The presence or absence of the remaining suspicious permissions was then considered as a feature. Our findings are shown in Figure 2.

I.Rassameeroj [14] states that certain permission combinations enable an application to perform dangerous actions posing threat to user's data and privacy. We considered these combinations as features for our feature set. Table 1 depicts the permissions and permission combinations considered as features.

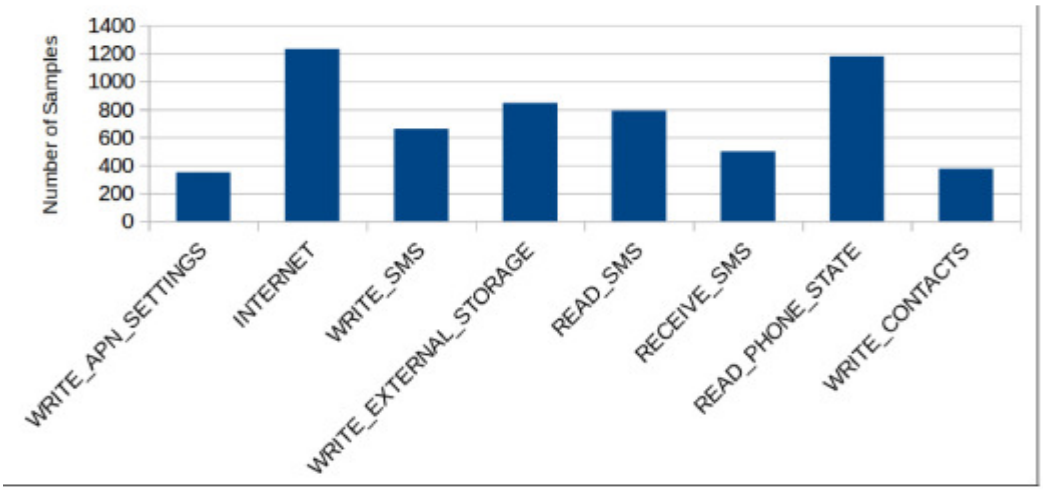


Figure 2. Frequency of suspicious permissions among malware samples

3.1.2. Suspicious API Combinations

APIs used by an application determines the actual functionality and capability of the application. Static analysis of APIs used in an application hence becomes important to understand what the application actually intends to do. In the similar direction of selecting permissions as features, our approach contributes by evaluating APIs extensively used by malware applications. APIs were broadly classified according to their usage by the application. From the list of APIs which are found in large number of malware samples, combinations were derived which could pose a threat to the user. Two main types of threats considered are financial losses and leakage of user's personal information. For example APIs for accessing user's personal information (network details, device ID, line number, etc.) in combination with APIs for sending SMS enables an application to transmit user's personal information to a predefined source. This leads to both breach of privacy as well as monetary loss. The monetary loss here is due to cost incurred when the SMS is sent.\par APIs for evaluation are extracted by disassembling classes.dex file using dexdump tool present in Android SDK [15]. Figure 3 depicts the a snapshot of classes.dex when disassembled using dexdump tool. Table 2 lists the API combinations considered as a feature for our feature set.

```
8582 018370: 0c09                                    |0010: move-result-object v9
8583 018372: 7100 d800 0008                          |0011: invoke-static {}, Landroid/telephony/SmsManager;.getDefault:()Landroid/telephony/
     SmsManager; // method@00d8
8584 018378: 0c0c                                    |0014: move-result-object v12
8585 01837a: 7210 5b00 0908                          |0015: invoke-interface {v9}, Landroid/database/Cursor;.getCount:()I // method@005b
8586 018380: 0a01                                    |0018: move-result v1
8587 018382: 3d01 0800                               |0019: if-lez v1, 0021 // +0008
8588 018386: 7210 5d00 0908                          |001b: invoke-interface {v9}, Landroid/database/Cursor;.moveToNext:()Z // method@005d
8589 01838c: 0a01                                    |001e: move-result v1
8590 01838e: 3901 0300                               |001f: if-nez v1, 0022 // +0003
8591 018392: 0e00                                    |0021: return-void
8592 018394: 1a01 b007                               |0022: const-string v1, "_id" // string@07b0
8593 018398: 7220 5a00 1908                          |0024: invoke-interface {v9, v1}, Landroid/database/Cursor;.getColumnIndex:(Ljava/lang/
     String;)I // method@085a
8594 01839e: 0a01                                    |0027: move-result v1
8595 0183a0: 7220 5c00 1908                          |0028: invoke-interface {v9, v1}, Landroid/database/Cursor;.getString:(I)Ljava/lang/
     String; // method@005c
8596 0183a6: 0c0a                                    |002b: move-result-object v10
8597 0183a8: 1a01 c80b                               |002c: const-string v1, "has_phone_number" // string@0bc8
8598 0183ac: 7220 5a00 1908                          |002e: invoke-interface {v9, v1}, Landroid/database/Cursor;.getColumnIndex:(Ljava/lang/
     String;)I // method@085a
```

Figure 3. Disassembled dex file

3.1.3. Manifest Violation

All the permissions required by an application should be declared in the AndroidManifest.xml. These permissions determine what are all the capabilities the application has. During application installation, all the permissions declared by the application are not cross verified by the package manager. Thus, at the run time if the application needs to perform a certain action and it does not have corresponding permission, run time exceptions occur. Malware developers take advantage of this flaw to perform collusion attacks [16]. The collusion attack requires at least 2 applications to work in collaboration. In this type of attack, an over privileged application provides an under privileged application with necessary permissions at runtime. Soundcomber [17] is one such application which aims at collecting user's information by capturing audio from device's microphone and then sends it over the network with help of another application having necessary permissions. Figure 4 depicts a scenario where two applications combine their permissions to read contacts and send them over the network.

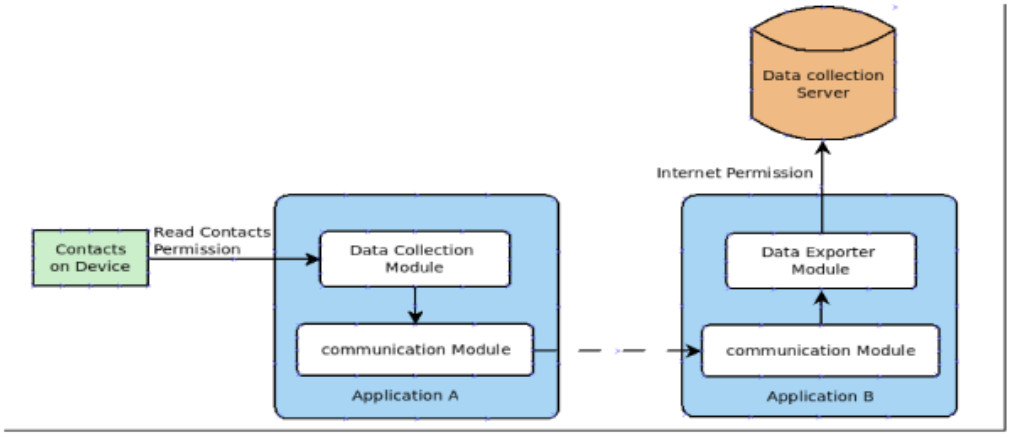


Figure 4. A collusion attack scenario

One way to detect the possibility of collusion attack is to look for application which has declared more permissions than what it requires (over privileged applications), but the drawback with this approach is the high false positive rate. The reason for high false positive rate is that many developers declare majority of the permissions available irrespective of their usage by the application.

We devised a different approach for detecting possible collusion attack. Rather than looking for over privileged applications we detect under privileged applications, that is the application declaring less permissions than what it actually required. The under privileged application then gets required privileges at runtime with the help of another application. To detect under privileged applications applications, we look for the permissions that will be used by the application at run time but are not present in application's manifest file. To derive permissions required by application at run time, permission required for executing each API present in application's dex file is extracted. If any permission required for execution of an API is not found in the application's manifest file, it is considered as a manifest violation.

We derive the permissions required by an API with the help of Android's developer guide and Pscout [18].

Each occurrence of manifest violation is assigned a weight of 7. A summation of these permission's weights was considered as the weight of the feature (Manifest violation).

3.1.4. Suspicious Content URI

A content URI (used for data access) can be called suspicious if by using that URI an application can leak user's personal data or can access another application's data. For example, an application can get access to contacts by using URI: content://com.Android.contacts. Such suspicious URIs were identified and their presence was checked among various malware and benign samples available in the training set. Suspicious content URIs which were detected in most of the malware samples and few benign samples were considered as a feature for feature set. Figure 5 shows the content URIs extensively used by malware applications.

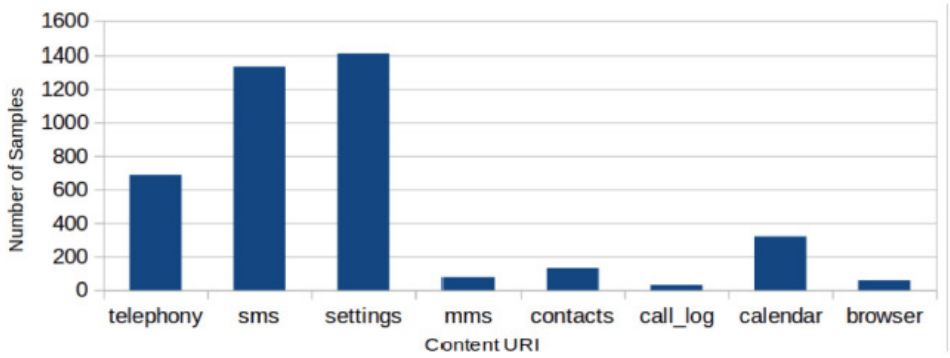


Figure 5. Frequency of suspicious content URIs among malware samples

To collect the content URIs used by the application, we parse the dalvik bytecode of disassembled classes.dex. The presence of content URIs that provide access to MMS, Browser and telephony data were seen among majority of malware applications.

Each Suspicious Content URI was assigned a weight of 6. Summation of the weights for frequency of such suspicious content URIs is considered as the weight of the feature.

3.1.5. Detection of Executable code

Embedding malicious code into documents has been successful technique for distributing malware. Desktop malware like Pidief, ZBOT, SillyD have been distributed as malicious PDF, JPEG, mp3 files. Based on Shafiq [19] and Stolfo's [20] findings which stated that detection of embedded malware requires parsing the bytecode of the documents, We employed a mechanism to find embedded executables by parsing the bytecode of all the files present in the resources directory of an APK. Many malware samples show the presence of executables and shell scripts embedded within image and music files. Presence of image files embedded with executable code can be found in samples from malware families like DroidKungFu1 and RougePush. Malware samples from DroidKungFu3 and GingerMaster families show presence of music files embedded with executable code.\par As this behavior was detected only in malware samples, presence of embedded executables was assigned a maximum weight of 10. Summation of the weights for frequency of such files is considered as weight of the feature.

3.2. Assigning Weight to Features

The weight assigned to a feature represents the impact that presence or absence feature makes on an application's classification. Weights are assigned to each feature on a scale of 1 to 10 using heuristics based approach such that higher the weight of a feature, more the feature contributes during classification. The highest weight of 10 was assigned to presence of executables embedded in image or music files. Presence of embedded executables is the strongest indicator in our feature set of an application being malicious as only malware samples are found to have resource files injected with executable code. All other features were assigned weights relative to the weight of 'presence of embedded executables' feature. Manifest violations are assigned a weight of 7. This is because unlike a malicious application, a benign application declares all the permissions being used. When compared to 'suspicious Permission combinations' or 'suspicious API combinations', 'manifest violation' has more impact during classification but it is not as influential as 'presence of embedded executables'. Thus it is assigned a weight lower than 'presence of embedded executables' and higher than 'suspicious Permission combinations' and 'suspicious API combinations'. Presence of suspicious content URI in an application is assigned a weight of 6. The presence of these content URI was seen in both malicious and benign samples, but number of malicious samples containing these URIs was much greater than number of benign samples. Weights for suspicious content URIs, manifest violations, presence of executable code are frequency based. Thus the total weight for these features in the feature set is multiple of the frequency of the feature occurrence and the weight assigned to the feature. \par Permission combinations and API combinations are assigned a moderate weight of 5 as the presence of these leads to suspicious behaviors, but their presence cannot conclude an application of being a malware or benign. We assigned suspicious permissions the lowest weight of 3 as these permissions can be found in large number in both benign and malware samples. Table 3 depicts the assignment of weights to the features selected.

3.3. Feature Vector Selection

After deciding upon the application's attributes to be considered as features, we considered and evaluated three categories of feature vectors with a set of machine learning algorithms. All the three categories of feature vectors constituted of similar features, but represented in different way. The first and second categories of feature vectors were weighted feature vector where as the third category was a non weighted feature vector. The first category of feature vector contained weights for each feature along with the Euclidean distance as an additional feature. The second category of feature vector was derived by excluding Euclidean distance from the first feature vector. For the third category of feature vector, rather than considering the frequency and weight of a feature, we check only presence of a feature. Representation in feature vector is done as either 1 or 0 to depict the presence or absence of a specific feature in the sample.

3.3.1. Evaluation of model for Feature Vector Selection

K-fold cross validation was carried out in order to evaluate the efficiency of the classification model. The default implementation of cross validation provided by WEKA was used for this purpose. The efficiency of the classifier models generated using all three categories of feature vectors were compared based on cross validation. One round of cross-validation of a two class classifier model involves segregating a sample of the training data set into two complementary

subsets, subset for performing the analysis (the training set) and subset for validating the analysis (the validation set). Inconsistency is reduced by multiple rounds of cross-validation using different segregations. Finally the average of all validation results is presented as true positive rate and false positive rate. We used WEKA [21] implementation for both model generation and cross validation. The true positive rate and false positive rate are deduced as follows :

$$TPR = \frac{TP}{TP + FN}$$

$$FPR = \frac{FP}{FP + TN}$$

Figure 6 (a) and Figure 6 (b) show variations in true positive rates and variations in false positive rates respectively for models generated using three categories of feature vectors.

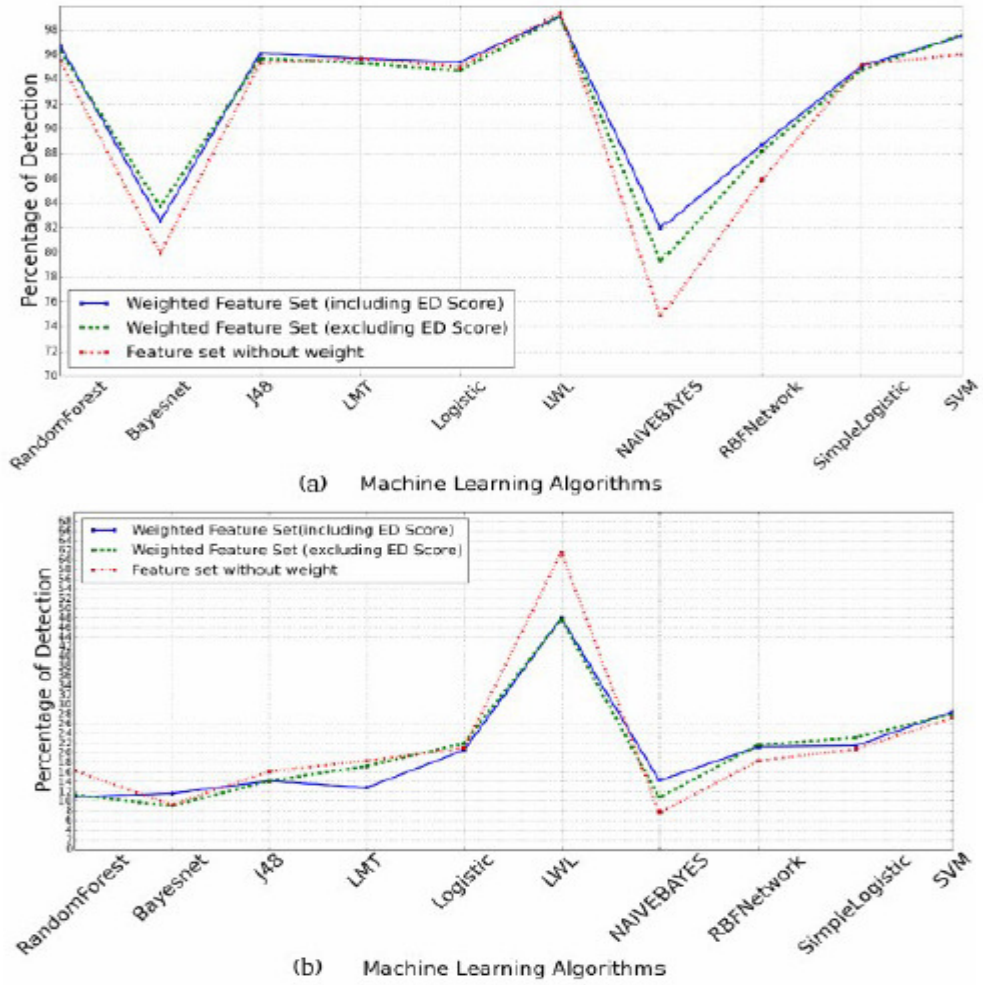


Figure 6. Variation inTPR (a) and FPR (b) for various models

High true positive and low false positive rates are observed for the second category of feature vector, that is a feature vector with weights and excluding Euclidean distance. Thus the second category of feature vector was considered for providing features to the machine learning algorithms. The reason for omitting Euclidean distance from the feature set was its last rank among the features on applying Chi-Square attribute ranking mechanism. This illustrated that excluding it as a feature would not affect the detection rates. Figure 7 shows variation in Euclidean distance across all the samples present in our dataset.

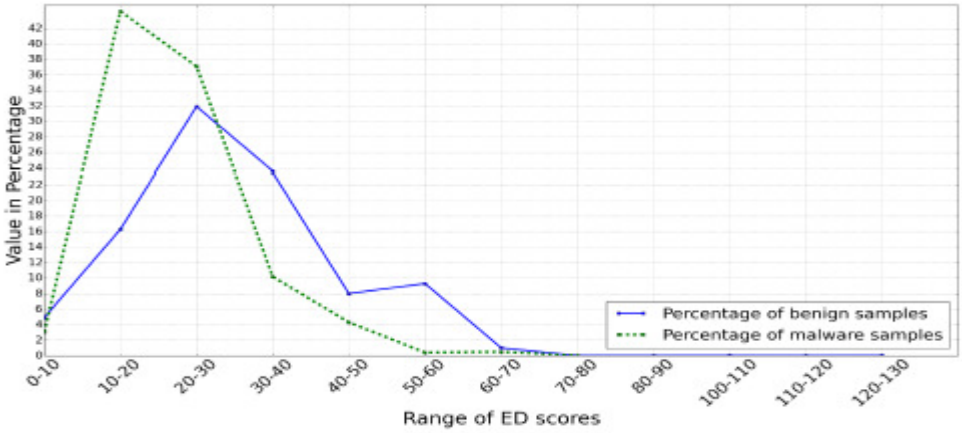


Figure 7. Variation in ED scores among benign and malware samples

Figure 8 shows the receiver operating characteristic (ROC) graph for the classification model built using second category of feature set. This graph illustrates the performance of a binary classifier system built using various machine learning algorithms and the weighted feature set. Random Forest algorithm depicts the maximum ROC space in the ROC curve which proves that for the given training set, classifier model built using Random Forest is more efficient than models generated using other machine learning algorithms. We used model built using Random Forest algorithm as the classifier in DroidSwan implementation.

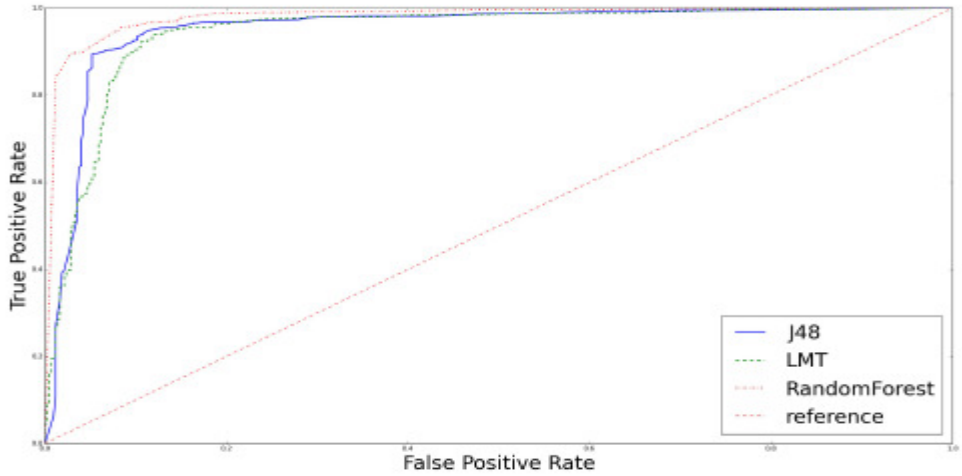


Figure 8. ROC Curve for classifier models based on various algorithms

3.4. DroidSwan working

Working of DroidSwan is carried out in two phases. Phase1 is the knowledge building phase. In this phase, DroidSwan extracts specific features and builds feature set of all the samples from the training set. These feature sets are then provided to the machine learning algorithm using WEKA implementation of machine learning algorithms. A two class classifier model is thus generated. \par Classifier model generated during phase 1 can be used for classification of samples without updating the model every time a new sample is provided for analysis.

Phase2 is the classification phase. In this phase, features are extracted from test application which needs to be classified and a corresponding feature set is built. Now this feature set is provided to the classification model generated during phase 1. The classification model then classifies the sample as either malicious or benign and generates a Json report for the same. This output report contains details regarding the presence or absence of all the features under consideration. The report also specifies all the suspicious content URIs and embedded executables present in the application. Figure 9 depicts DroidSwan's architecture.

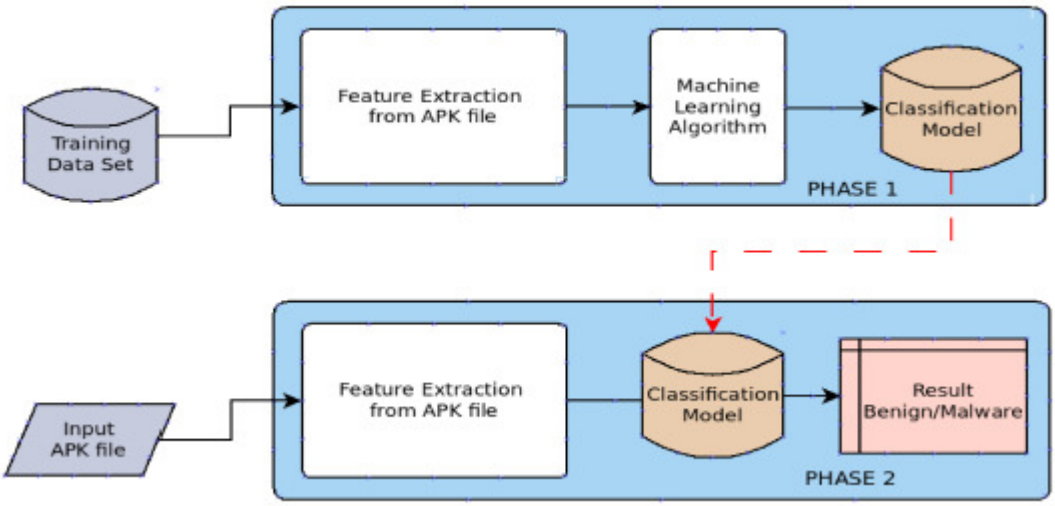


Figure 9. DroidSwan architecture

4. RESULTS

The efficiency of DrooidSwan classification model was tested by analyzing 500 malware samples obtained from Virustotal malware intelligence service \cite{hispasec2011virustotal} and 800 benign samples from ApkDrawer \cite{apkDrawer}. Collectively these samples constituted of our test-set. It was verified beforehand that the test-set does not contain any samples in common with the training-set by comparing the hashcode of each sample in test set against hashcodes of samples from training set. Figure 10 and Figure 11 depict the detection rates of malware samples and benign samples respectively by using DroidSwan.

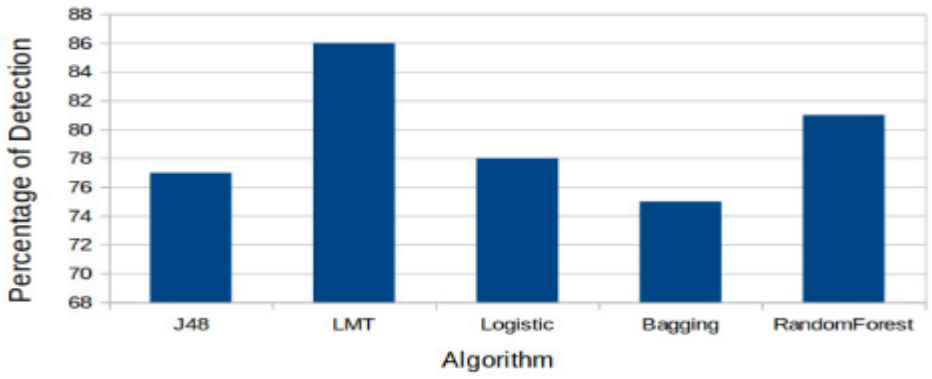


Figure 10. Detection rate of DroiidSwan for malware samples

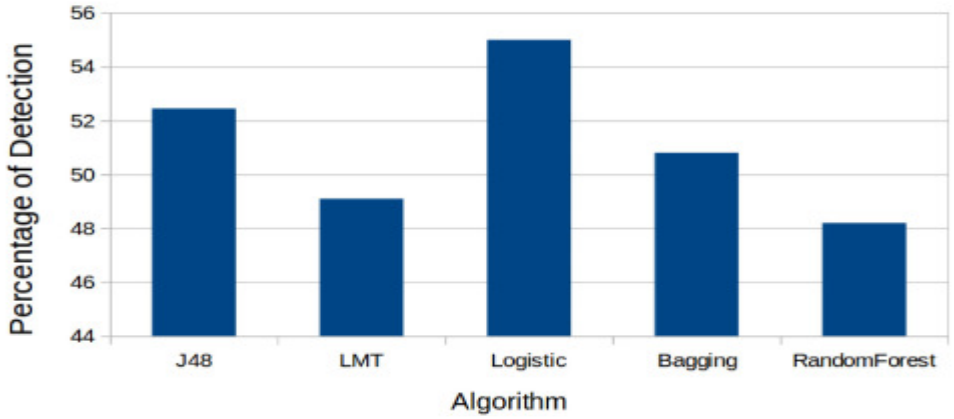


Figure 11. Detection rate of DroiidSwan for benign samples

The detection rate of DroidSwan was compared with the detection rates of four other antivirus solutions for the same set of malware samples. Figure 12 shows the detection rate of DroidSwan in comparison with Kaspersky (version 12.0.0.1225) [23], McAfee (version 6.0.5.614) [24], Avast (version 8.0.1489.320) [25] and TrendMicro (version 9.740.0.1012) [26].

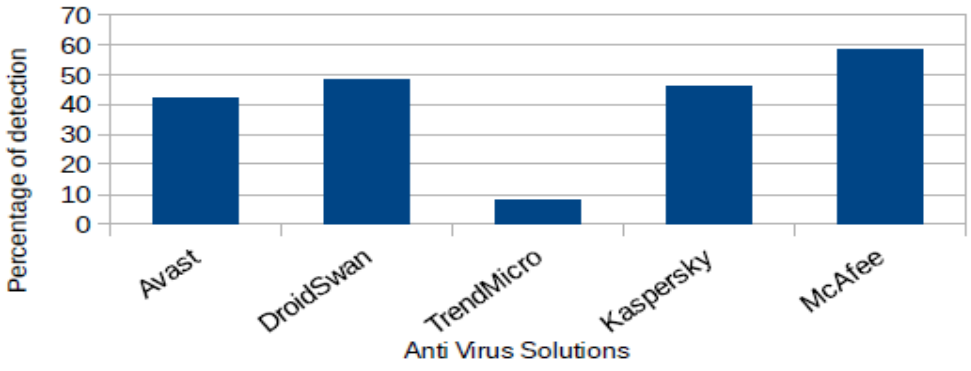


Figure 12. Detection rates of DroiidSwan in comparison with other AV solutions

Recall rate of DroidSwan with Random Forest based classifier for malwares from various malware families is shown in Figure 13.

5. CONCLUSION

We present DroidSwan, an approach for detecting malicious Android applications wholly based on static analysis of their respective APK files. The process of classification comprises of extracting 24 features, assigning weights to the features and finally using the collection of feature weights as a feature set. The feature set along with Random Forest classifier model is then used to classify the given sample as either malware or benign.\par We observed that classifier model built using Random Forest shows higher TPR and lower FPR when compared to other machine learning algorithms.

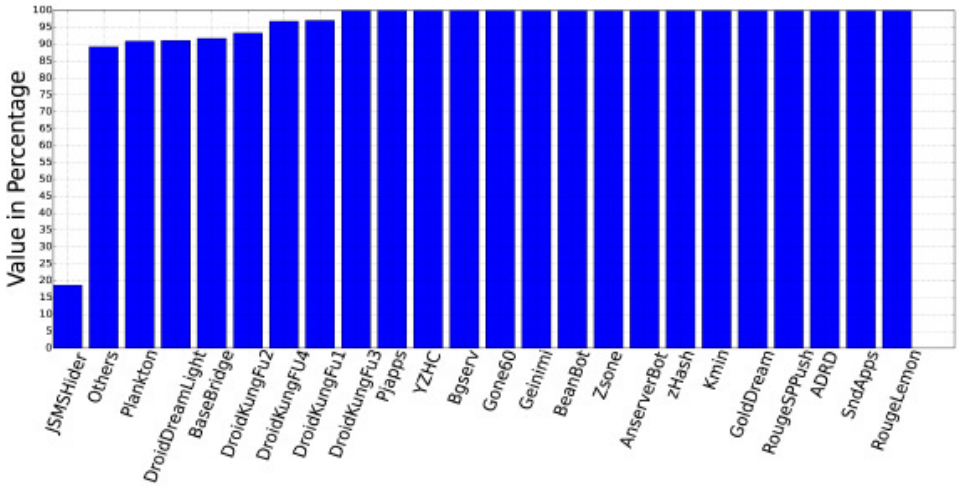


Figure 13. Recall rate of DroidSwan for various malware families

REFERENCES

[1] RiskIQ, Feb 19 2014, Research Also Shows Steady and Significant Drop in Number of Malicious Apps Being Removed in Past Three Years. Available: http://www.riskiq.com/company/press-releases/riskiqreports-malicious-mobile-apps-google-play-have-spiked-nearly-400

[2] Genome Project. Android malware samples. http://www.malgenomeproject.org.

[3] S. Hispasec Sistemas. Virustotal malware intelligence service, 2011.

[4] J. Erdfelt. Apkparser tool. https://code.google.com/p/xml-apk-parser.

[5] A. Desnos. Androguard. Available at https://code.google.com/p/androguard/.

[6] Schmidt, A-D., Rainer Bye, H-G. Schmidt, Jan Clausen, Osman Kiraz, Kamer A. Yuksel, Seyit Ahmet Camtepe, and Sahin Albayrak. "Static analysis of executables for collaborative malware detection on android." In Communications, 2009. ICC'09. IEEE International Conference on, pp. 1-5. IEEE, 2009.

[7] Zhou, Yajin, Zhi Wang, Wu Zhou, and Xuxian Jiang. "Hey, You, Get Off of My Market: Detecting Malicious Apps in Official and Alternative Android Markets." In NDSS. 2012.

[8] Wu, Dong-Jie, Ching-Hao Mao, Te-En Wei, Hahn-Ming Lee, and KuoPing Wu. "Droidmat: Android malware detection through manifest and API calls tracing."In Information Security (Asia JCIS), 2012 Seventh Asia Joint Conference on, pp. 62-69. IEEE, 2012.

[9] Felt, Adrienne Porter, et al. ”Android permissions demystified.”Proceedings of the 18th ACM conference on Computer and communications security. ACM, 2011.
[10] Enck William, Machigar Ongtang, and Patrick McDaniel. ”On lightweight mobile phone application certification.” Proceedings of the 16th ACM conference on Computer and communications security. ACM, 2009.
[11] Zheng, Min, Mingshen Sun, and John Lui. ”Droid Analytics: A Signature Based Analytic System to Collect, Extract, Analyze and Associate Android Malware.” Trust, Security and Privacy in Computing and Communications (TrustCom), 2013 12th IEEE International Conference on IEEE, 2013.
[12] Shabtai, Asaf, Yuval Fledel, and Yuval Elovici. "Automated static code analysis for classifying Android applications using machine learning." Computational Intelligence and Security (CIS), 2010 International Conference on. IEEE, 2010.
[13] Zhou, Yajin, and Xuxian Jiang. "Dissecting android malware: Characterization and evolution." Security and Privacy (SP), 2012 IEEE Symposium on. IEEE, 2012.
[14] Rassameeroj, Ittipon, and Yuzuru Tanahashi. "Various approaches in analyzing Android applications with its permission-based security models." Electro/Information Technology (EIT), 2011 IEEE International Conference on. IEEE, 2011.
[15] Google Inc. Official Page for android developers. http://developer.android.com.
[16] Bugiel, Sven, Lucas Davi, Alexandra Dmitrienko, Thomas Fischer, Ahmad-Reza Sadeghi, and Bhargava Shastry. "Towards Taming Privilege-Escalation Attacks on Android." In NDSS. 2012.
[17] Schlegel, Roman and Zhang, Kehuan and Zhou, Xiao-yong and Intwala, Mehool and Kapadia, Apu and Wang, XiaoFeng. 'Soundcomber: A Stealthy and Context-Aware Sound Trojan for Smartphones.'NDSS, 2011
[18] Au, Kathy Wain Yee, Yi Fan Zhou, Zhen Huang, and David Lie. "Pscout: analyzing the android permission specification." In Proceedings of the 2012 ACM conference on Computer and communications security, pp. 217-228. ACM, 2012.
[19] Shafiq, M. Zubair, Syed Ali Khayam, and Muddassar Farooq. "Embedded malware detection using markov n-grams." In Detection of Intrusions and Malware, and Vulnerability Assessment, pp. 88-107. Springer Berlin Heidelberg, 2008.
[20] Stolfo, Salvatore J., Ke Wang, and Wei-Jen Li. ”Towards stealthy malware detection.” Malware Detection. Springer US, 2007. 231-249.
[21] Hall, Mark, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. ”The WEKA data mining software: an update.” ACM SIGKDD explorations newsletter 11, no. 1 (2009): 10-18.
[22] Z. Jay. Apkdrawer.com. http://www.apkdrawer.com.
[23] Kaspersky mobile security. Available at http://www.kaspersky.co.in/downloads/android-security.
[24] Mcafee mobile security. Available at https://www.mcafeemobilesecurity.com/.
[25] Avast mobile security. Available at http://www.avast.com/en-in/free-mobile-security.
[26] Trendmicro mobile security. Available at http://www.trendmicro.com/us/enterprise/product-security/mobile-security/.

APPENDIX

Table 1. Suspicious permissions and permission combinations

Suspicious permissions and permission combinations	Weight assigned
READ SMS	3
WRITE SMS	3
RECEIVE SMS	3
WRITE CONTACTS	3
WRITE APN SETTINGS	3
SEND SMS	3
ONLY INTERNET	3
ONLY WRITE EXTERNAL STORAGE	3
WRITE SMS and RECEIVE SMS	5
SEND SMS and WRITE SMS	5
INTERNET and WRITE EXTERNAL STORAGE	5
INTERNET,RECORD AUDIO, READ PHONE STATEand MODIFY PHONE STATE	5
ACCESS FINE LOCATION or ACCESS COARSE LOCATION, RECEIVE BOOT COMPLETED and INTERNET	5
INTERNET,RECORD AUDIO and PROCESS OUTGOING CALLS	5

Table 2. Suspicious API combinations

Suspicious API combinations	Weight assigned
"landroid/telephony/telephonymanager;.getdeviceid" "landroid/location/locationmanager;.getlastknownlocation" "landroid/location/location;.getlatitude" "landroid/location/location;.getlongitude" "landroid/telephony/smsmanager;.sendtextmessage" "landroid/net/uri;.parse","landroid/location/locationmanager;.getbestprovider"	5
"ljava/net/urlencoder;.encode" "ljava/net/uri;.getquery" "ljava/net/httpurlconnection;.connect" "ljava/net/httpurlconnection;.geturl" "ljava/net/httpurlconnection;.getheaderfield" "landroid/location/locationmanager;.getbestprovider" "landroid/location/location;.getlatitude" "landroid/location/location;.getlongitude" "landroid/telephony/gsm/smsmanager;.sendtextmessage"	5
"landroid/net/uri;.parse" "landroid/content/contentresolver;.query" "landroid/database/cursor;.movetonext" "landroid/database/cursor;.getcolumnindex"	5

"landroid/database/cursor;.getstring" "landroid/database/cursor;.close" "landroid/database/cursor;.movetolast" "landroid/database/cursor;.movetoprevious"	
"landroid/net/uri;.parse" "ljava/net/urlencoder;.encode" "ljava/net/url;.openstream" "landroid/telephony/telephonymanager;.getdeviceid" "landroid/telephony/telephonymanager;.getline1number" "landroid/telephony/telephonymanager;.getnetworkcountryiso" "landroid/telephony/telephonymanager;.getnetworkoperatorname" "ljava/io/bufferedreader;.readline" "landroid/content/pm/packagemanager;.hassystemfeature"	5
"ljava/net/inetaddress;.getlocalhost" "ljava/net/inetaddress;.gethostname" "ljava/net/url;.openstream" "ljava/net/inetaddress;.getbyname" "ljava/net/inetaddress;.equals" "ljava/net/inetaddress;.hashcode" "landroid/net/uri;.parse" "landroid/telephony/smsmanager;.getdefault" "landroid/telephony/smsmanager;.dividemessage" "landroid/telephony/smsmanager;.sendtextmessage" "landroid/telephony/telephonymanager;.getdeviceid" "landroid/telephony/telephonymanager;.listen"	5
"ljava/net/urlencoder;.encode" "ljava/net/uri;.<init>" "landroid/location/location;.hasaccuracy" "landroid/location/location;.distanceto" "landroid/location/location;.gettime" "landroid/location/location;.getaccuracy" "landroid/location/location;.getlatitude" "landroid/location/location;.getlongitude" "landroid/location/location;.getprovider" "landroid/location/locationmanager;.requestlocationupdates" "landroid/location/location;.<init>" "landroid/location/location;.setaccuracy"	5
"ljava/net/urlencoder;.encode" "ljava/net/url;.<init>" "ljava/net/url;.openconnection" "landroid/telephony/telephonymanager;.getline1number" "landroid/telephony/smsmanager;.getdefault" "landroid/telephony/smsmanager;.sendtextmessage" "landroid/telephony/smsmessage;.getdisplayoriginatingaddress" "landroid/telephony/smsmessage;.getmessagebody" "landroid/telephony/smsmessage;.createfrompdu"	5

Table 3. Weight assignment to various features

Feature type	Weight assigned
Suspicious permissions	3
Suspicious permission combinations	5
Suspicious API combinations	5
Suspicious content URI	6
Manifest violation	7
Presence of executable	10

AUTHORS

Babu Rajesh V has been working for three years in the field of mobile security and malware analysis. His areas of interests include mobile security and embedded security.

Phaninder Reddy has been working for two years in the field of mobile security and malware analysis. His areas of interests include machine learning and data analytics.

Himanshu Pareek has around six years of experience in developing and design of security solutions related to small sized networks. He has research papers published on topics like malware detection based on behavior and application modeling.

Mahesh U Patil received master degree in electronics and communication. Presently he is working as Principal Technical Officer at Centre for Development of Advanced Computing. His research interests include Mobile Security and Embedded Systems.

EMBEDDING AND EXTRACTION TECHNIQUES FOR MEDICAL IMAGES – ISSUES AND CHALLENGES

S.Priya[1] and R.Varatharajan[2]

[1]Research Scholar/ECE, Bharath University , Chennai
priyarakaesh@gmail.com
[2]Professor and Head,Department of ECE,
Sri Lakshmi Ammal Engineering College, Chennai
varathu21@yahoo.com

ABSTRACT

New technologies in multimedia and communication fields have introduced new ways to transfer and save the medical image data through open networks, which has introduced new risks of inappropriate use of medical information. Medical images are highly sensitive hence secured transmission and reception of data is needed with minimal distortion. Medical image security plays an important role in the field of Telemedicine. Telemedicine has numerous applications in teleconsulting, teleradiology, telediagnosis, telesurgery and remote medical education. Our work is to analyze about the different embedding techniques that can be used for embedding the personal and diagnosed details of a person within the medical images without any visual discrepancy. Also to survey about the blind extraction algorithm utilizing genetic algorithm for optimization of the key parameters.

KEYWORDS

Robustness, Fidelity, Joint Watermarking, Embedding algorithm, Genetic algorithm

1. INTRODUCTION

In the new era medical images are produced from a variety of imaging equipments, such as Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), Single Photon Emission Computed Tomography (SPECT), etc. CT can clearly reflect the anatomical structure of bone tissues. SPECT can highlight the lesion of tissues and organs to provide information about blood flow and temperature of body parts. PET scanning can show blood flow, oxygen and glucose metabolism in the tissues of the brain. MRI can clearly reflect the anatomical structure of soft tissues, organs and blood vessels. These medical images can also be saved in a digital format.

Electronic Patient Record (EPR) is one of the digital format of saving the personal details of patient, details of diagnosis, hospital information along with the medical image for continual monitoring [1]. Telemedicine is the process by which electronic, visual and audio

David C. Wyld et al. (Eds) : ACITY, DPPR, VLSI, WiMNET, AIAA, CNDC - 2015
pp. 179–192, 2015. © CS & IT-CSCP 2015 DOI : 10.5121/csit.2015.51316

communications are used to support practitioners at remote sites with diagnosis and consultation procedures, such as remote clinical examinations and medical image transfers. Telemedicine is legally regulated by laws and constraints regarding the access of data contained in Personal medical Files. The transmit of medical transcriptions through online provides efficient clinical interpretation without carrying the documents. The diagnosis needs confidentiality, availability, and reliability [2]. Confidentiality means that only the original users have access to the information. Availability, guarantees access to medical information. Reliability is based on integrity that the information has not been modified by unauthorized persons; and authentication intends that the information belongs indeed to the correct patient.

Hospital patient database management system is implemented in more hospitals nowadays which is designed to transform the manual way of maintaining and accessing patient medical files into electronic medical record (EMR) or Electronic Health Record (EHR) [3] . EMR is used to solve the problem of manual method. To protect this private information about a person against unauthorized viewers, we can use any one of the techniques for medical image data : Cryptography or Watermarking.

Cryptography is the technique of transforming information to more secured form. Digital encryption of medical images before transmission and storage is proposed as an easiest way to protect the patient information. Cryptography technique can be divided into symmetric encryption needs secret key and asymmetric encryption which needs private and public keys. An encryption technique [4] is used which is a blend of symmetric key encryption and steganography with a variable length key derived from the encrypted text itself to have better security. Chaotic systems [5] [6] can be used for medical images to achieve robust system and its implemented using Bit Recirculation Image Encryption (BRIE) to control the pseudo-random operations on each pixels. The traditional algorithms are not recommended for medical images because they are slow, hence we can associate the properties of traditional cryptography with the properties of a chaotic system. There are several chaotic algorithms which handle the medical image are Bit Recirculation Image Encryption, Circulation Encryption Algorithm, Chaotic Key-Based algorithm, Image Chaotique, Hierarchique Encryption. Visual cryptographic technique [7] can also be used for encrypting medical images which provides a more reliable system.

The cryptography systems have only limited techniques hence the tracker can easily predict the original image after some iterations .Hence we should choose a technique where the predictions of the data from the embedded data should be so difficult by the unauthorized users. Watermarking has found a niche role in secure sharing and handling of medical images. The watermarks are embedded into medical image for three purposes: hiding electronic patient's data, integrity verification and for authentication.

The watermarking techniques are divided into two basic categories as spatial domain watermarking and Frequency domain watermarking. In spatial domain the Least Significant Bit (LSB) of the image pixel is replaced with the watermark bit and in the Frequency domain the image is transformed to the frequency domain and then the frequency components are modified with the watermark bits.

The watermarking techniques can also be classified based on the watermarking robustness as Robust, Fragile and Semi-Fragile [8] [9]. Robust watermarks can resist non-malicious distortions and best suited for copyright protection Fragile watermarks can easily destroyed by all image

distortions and its suited for tamper detection and authentication. Semi-Fragile watermarks can be destroyed by certain types of distortions and resists minor changes and used for some special cases of authentication.

There are three watermark detection/extraction schemes: Non-blind, semi-blind, and blind. Both the original image and the secret key are needed for non-blind extraction . Semi-blind needs only the secret key and the watermark. Blind extraction system needs only the secret key. The digital watermark when it is hidden in the image it generally introduces some amount of imperceptible distortion in the image. In medical images, there is a region that is important for diagnosis called ROI (region of interest) and RONI (region of non-interest) . Embedding data in ROI region should not cause any visual artifacts which affects the interpretation by medical doctors. So watermarking can be used in RONI of medical image [11].To enhance confidentiality and authentication a dual watermarking scheme in which Caption watermarking for hiding patient's information in ROI and Signature watermarking hides the physician's digital signature in RONI .

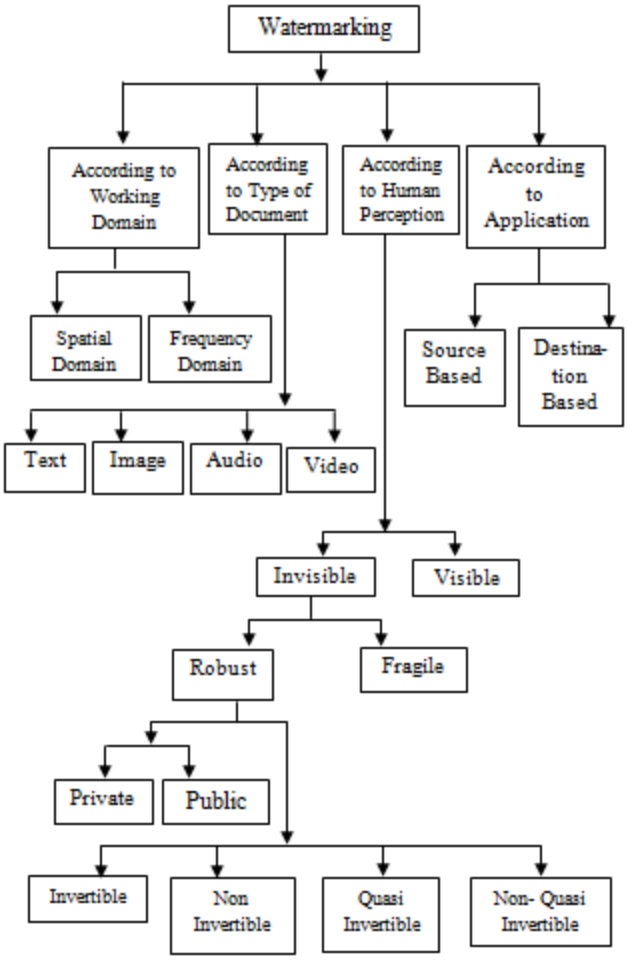


Fig 1. Classification of Watermarking

To achieve better performance in terms of perceptually, invisibility and robustness, an adaptive quantization parameters can be used for data hiding. The embedding strength is more or less proportional to the value of energy to have better robustness and transparency. The block wise embedding technique with the larger quantization parameters improve the robustness. Another dual watermarking method consists of an annotation part and a fragile part [11] in which Encrypted patient data is embedded in an annotation watermark, and tampering can be detected using a fragile watermark.

In order for physicians, hospitals, and patients to utilize the benefits of digital medical images, all communication must be compatible. When manufacturers use proprietary formats, the digital files can be read only with the manufacturer's equipments and communicating these files over multiple networks is not possible.As digital medical information evolved, the medical community demands for a standard method of transmitting medical images and their associated information. DICOM is a standard file format for transmission and storage of digital medical images which suits for hospital database management system [12] [13]. DICOM defines the network and media interchange services allowing consistency so that EHR records are available to all who need them.

Header in DICOM image format stores patient's information such as patient identification number, name, sex, and age. Insurance companies, hospitals and patients may want to change this data for various reasons.. After embedding the data, watermarked medical image can still conform to the DICOM format. Pixel data can be compressed using a compression standard and the DICOM Grayscale Display Function Standard improves image quality. The DICOM Presentation State Storage Service Class ensures presentation consistency since the physician interpreting an image may adjust magnification, window width, and window center or apply various image processing enhancements. DICOM has a key role in virtually every medical profession that uses images, including cardiology, dentistry, endoscopy, mammography, ophthalmology, orthopedics, pathology, pediatrics, radiation therapy, radiology, and surgery as well as veterinary medical imaging. DICOM continuously updates the standards per year.

2. RELATED WORKS

2.1. Watermarking Techniques:

a. Spatial Domain Techniques:

In Spatial domain the watermark is directly embedded by modifying the pixels of the original image without any transformation of the image. This technique is often fragile and applied in the pixel domain and has less complex computation thus consumes less time for archiving and retrieval. The least significant bit (LSB) technique is used to embed information [14] in a cover image. The LSB technique of a cover image is described by changing pixels by bits of the secret message. An embedding scheme which randomly hides messages in the LSB of any/all component of the chosen pixel using polynomial [15]. If polynomial is used, hacker needs to predict more than one number i.e. all coefficients of polynomial has to be decoded correctly and probability of finding all right coefficients is less compared to predicting single bit. Watermarking can be done by embedding watermark into sub images with LSB technique. The watermark can be embedded into specifics blocks[16] of the host image where the selection of blocks are based on entropy value which gives a high PSNR value.

b. Frequency Domain Techniques:

Transformation of an image is needed to get more information about the image and to reduce the computational complexity. Even though this technique takes more time and more complex than spatial domain technique the embedded watermarked data cannot be identified easily as the previous technique. In transform domain the watermark is embedded after performing transformations such as, Discrete Cosine Transform (DCT), Discrete Fourier Transform (DFT), and Discrete Wavelet Transform (DWT), Contourlet Transform etc. The watermark is embedded in the transform coefficients. When compared to spatial domain these techniques offer high security and are robust to attacks. In frequency domain watermarking the values of selected frequencies can be altered. Since high frequencies will be lost by compression or scaling, the watermark signal is applied to lower frequencies, or better yet, applied adaptively to frequencies containing important elements of the original picture.

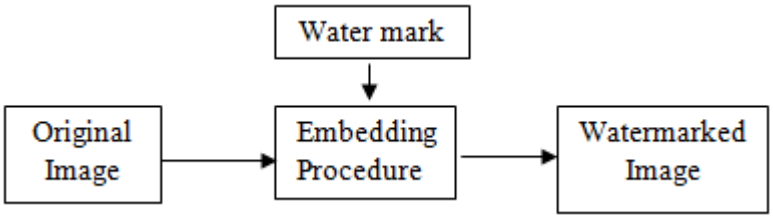


Fig 2. Embedding Process

DFT is the basic transform used for data and images.DCT is a fast transformation technique provides excellent energy compaction for highly correlated data and most of the information (dc-coefficient) is in the first pixel. DWT gives both the frequency and location in time and it is suited for Time-varying signals. The Contourlet Transform has the main feature of capturing two dimensional singularities. With an intention to increase the robustness a reversible technique based on wavelet transform is proposed by Lingling An et al. [17]. SQH with k-means clustering is used to resist unintentional attacks and EPWM is used to balance invisibility and robustness.

A new approach of Lifting Wavelet Method (LWT) is capable of maintaining the structural index and provides a better performance than the DWT method. The Daubechies wavelet family performs better than the other wavelet types. A dual security approach can be employed for medical image security where the medical image is considered as watermark and is watermarked inside a natural image. This approach is to wean way the potential attacker by disguising the medical image as a natural image. To enhance the security the watermarked image can be encrypted using encryption algorithms. The water marking can be implemented using Lifted Wavelet Transforms (LWT) and Singular Value Decomposition (SVD) technique [18].The features of Lifting Wavelet Transforms (LWT) along with Discrete Wavelet Transforms (DWT) and Singular Value Decomposition (SVD) [19] can be used to provide a robust and imperceptible watermark. Contourlet Transform provides a multi resolution and directional expansion of images using Pyramidal Directional Filter Bank (PDFB). The LP decomposes the image into frequency band to obtain singular points. he DFB decomposes each LP detail band to capture directionality.

C. Joint Watermarking:

Joint watermarking combines the watermarking of the encrypted data in medical images in order to provide more security utilizing the benefits of cryptography and watermarking techniques.

Joint Medical image watermarking is to encapsulate vital data inside an image in energy packed areas, which is optimized with respect to image quality and to provide a second level security by incorporation of state of the art cryptographic standard. Rajendra Acharya et al [20] has proposed the technique of watermarking by interleaving encrypted patient information with medical images during JPEG compression, to reduce storage and transmission overheads. A novel method for watermarking by SVD based blind watermarking method and ciphering color in ages, based on the joint use of a key-dependent wavelet transform ,Fibonacci-Haar wavelet transform domain to increase its security with a secure cryptographic scheme by Federica Battisti et al [21]. Gouenou Coatrieux et al [22] has proposed a new technique of knowledge digest which gives a synthetic description of the image content, a digest that can be used for retrieving similar images with either the same findings or differential diagnoses. A secure version of a classical trellis coded quantization watermarking where the trellis path generated from the discrete key and the message was given by Sofiane Braci et al [23].The spread transform can represent a second or alternative security level for watermarking systems which makes message hard to read for unauthorised user. The access control model in order to enhance the protection of medical images was given by Wei Pan et al [24] in distributed healthcare infrastructures. Mohammad-Saleh Nambakhsha et al [25] has used digital watermarking framework using electrocardiograph (ECG) and demographic text data as double watermarks. The watermarks are embedded in selected texture regions of a PET image using multi-resolution wavelet decomposition.

A secure watermarking system using Arnold scrambling and 2-D Cellular Automata Transform (CAT) was given by Xiao-Wei Li et al.[26] . CAT-based watermarking system can simultaneously improve security, robustness and image quality of the watermarked image. Dalel Bouslimia,B et al [27] has proposed a joint encryption/watermarking algorithm in which it combines the RC4 stream cipher and two substitutive watermarking modulations: the Least Significant Bit Method and the Quantization Index Modulation which improves the peak signal to noise ratio . Dalel Bouslimi et al [28] a joint encryption/watermarking system based on an approach which combines a substitutive watermarking algorithm, the quantization index modulation, with an encryption algorithm: a stream cipher algorithm (e.g., the RC4) or a block cipher algorithm (e.g., the AES in cipher block chaining (CBC) mode of operation). The algorithm of discrete wavelet transform and Hankel transform combined is developed to achieve the integrity authentication of color image contents through embedding watermarking by M.V.S.S.Babu et al [29]. Vinay Pandey et al [30] has used steganography by medical image of any other as cover image and embedded encrypted image as secrete image with the private key.It also apply two shares encryption algorithm for encryption of embedded image. To achieve integrity service Mohamed M. Abd-Eldayem et al [31] has proposed a a hash value based encryption.To provide confidentiality and authentication services: the compressed R–S-Vector, the hash value and patient ID are concatenated to form a watermark then this watermark is encrypted using AES encryption technique. Hung-I Hsiao et al [32] has used chaotic amplitude phase frequency model (APFM) nonlinear adaptive filter for medical image security using is proposed. We set nine parameters, simulated time interval, and initial values for APFM nonlinear adaptive filter to generate chaotic orbits. Lamri Laouamer et al [33] has proposed a new approach for generating symmetric keys for image encryption / decryption, whereby the medical images (area of interest) use an informed process based on a technique that has been demonstrated on textual analysis called N-grams. Watermarking is performed by using a new nontensor product wavelet filter banks designed by A. Kannammal et al [34], which have the ability to reveal singularities in different directions. Natural image is taken as the original image

and the medical image is taken as a watermark image. The proposed algorithm has the ability to withstand different attacks like noise, rotation, contrast, and brightness attacks.

A Joint FED watermarking system is proposed by . P. Viswanathan et al [35] for addressing the issues of teleradiology. The system combines a region based substitution dual watermarking algorithm using spatial fusion, stream cipher algorithm using symmetric key, and fingerprint verification algorithm using invariants. Medical information, is embedded into the regions of interest (ROI) in medical images with a high capacity difference-histogrambased reversible data-hiding scheme. After that, the watermarked medical images are encrypted with hyperchaotic systems proposed by Shun Zhange et al [36].V. Amutha et al [37] has proposed a substitutive watermarking algorithm combined with an encryption algorithm, advanced encryption standard (AES) in counter mode.

3. BLIND WATERMARKING

It is a Zero-Knowledge watermarking algorithm which does need the original image for the detection process. Wei-Hung Lin A [38] has proposed a blind watermarking algorithm based on maximum wavelet coefficient quantization for copyright protection. The watermark is embedded in the local maximum coefficient which can effectively resist attacks. either non-geometry or geometry attacks. The watermark can effectively resist common image processing attacks, especially by JPEG compression (with a quality factor greater than 20) and Gaussian noise with a variation of less than 2. Xinge You [39] proposed a new method for constructing nontensor product wavelet filter banks and applied them into watermarking scheme design. The proposed wavelet filter banks make the watermarking scheme more flexible because more subbands and coefficients are suitable for watermark embedding. The algorithm is robust against various attacks particularly Gaussian noising attack.

Singular value decomposition (SVD) based image watermarking technique was proposed by Deepa Mathew K [40]. SVD uses non fixed orthogonal bases. The result of SVD gives good accuracy, good robustness and good imperceptibility. Swanirbhar Majumder [41] has used singular value decomposition (SVD) with the unconventional transform called Contourlet transform (CT) . Here the combination of pyramidal and directional filter bank (PDFB) has been used in Contourlet transform. An effective watermarking algorithm based on the chaotic maps was proposed by Jila Ayubi [42]. The chaotic maps are employed to generate a key space with the length of 10^{40} numbers to increase the degree of security. Mutation operator has been used to encrypt the watermark. This algorithm can preserve the hidden information against geometric and non geometric attacks. Surya Pratab Singh [43] has used a 3rd level of DWT (Discrete Wavelet transform) and before embedding the watermark image is passed through chaotic encryption process for its security, Other important thing is that in this watermark is embedded in the form of DCT (Discrete Cosine Transform) with special coefficient shifting algorithm to minimize the impact on main image. Yinglan Fang [44] has proposed an improved blind watermarking algorithm based on two-dimensional discrete wavelet transform. Before embedded watermark, the watermarking image is pretreated by using Arnold scrambling to improve its security. This algorithm had better concealment and improved the robustness and efficiency. A new blind watermarking algorithm technique based on DCT and DWT using middle frequency band of DCT and 2- levels DWT by Farhed aseed [45]. Using ARNOLD transformation the robustness and security of watermark image is increased. Nidhi Bisla [46] has used the watermarking technique of DWT and hybrid DWT-SVD. In case of DWT, decomposition of the original

image is done to embed the watermark and in case of hybrid DWT-SVD firstly image is decomposed according to DWT and then watermark is embedded in singular values obtained by applying SVD. Sudeb Das [47] has proposed a blind, fragile and Region of Interest (ROI) lossless medical image watermarking (MIW) technique, providing an all-in-one solution tool to various medical data distribution and management issues like security, content authentication, safe archiving, controlled access retrieval, and captioning.

A novel image encryption technique was given by J.B.Lima A [48] which involves two steps, where the finite field cosine transform is recursively applied to blocks of a given image. In the first step, the image blocks to be transformed result from the regular partition of subimages of the original image. The transformed subimages are regrouped and an intermediate image is constructed. In the second step, a secret-key determines the positions of the intermediate image blocks to be transformed. N. Venkatram [49] has highlights the extension of dwt-svd based image watermarking to medical images. For medical images 2D lifting wavelet transform (LWT) is used instead of dyadic 2D discrete wavelet transform. Even after attacks LWT-SVD method gives satisfactory quality both visually and mathematically.B. Jagadeesh [50] has used a novel digital image watermarking algorithm based on artificial neural networks. As neural networks are good at pattern recognition, they can be used as a medium to store the frequency domain components of the image and these can be used at the extraction of watermark. It gives better PSNR value and it is robust to many image processing attacks like compression, resizing & filtering.

4. GENETIC ALGORITHM

Genetic algorithms are adaptive algorithms for finding the global optimum solution for an optimization problem. It is also called as an optimization algorithm, meaning they are used to find the optimal solution(s) to a given computational problem that maximizes or minimizes a particular function. These algorithms are far more powerful and efficient than random search and exhaustive search algorithms, yet require no extra information about the given problem. This feature allows them to find solutions to problems that other optimization methods cannot handle due to a lack of continuity, derivatives, linearity, or other features. GAs applying the principles of survival of the fittest , selection , reproduction , crossover (recombining) , and mutation on these individuals to get , hopefully , a new butter individuals (new solutions) . There are two basic genetic algorithms operators which are crossover and mutation. These two operators are work together to explore and exploit the search space by creating new variants in the chromosomes. Researchers used Genetic Algorithm to optimize the watermarking requirements.

Hamed Modaghegh et al [51] used a new adjustable watermarking method based on singular value decomposition is presented so that SVD parameters are adjusted by using the GA considering image complexity and attack resistance. Veysel Aslantaset al [52] has used intelligent optimization algorithms for correcting the rounding errors caused by the transformation process. K. Ramanjaneyulu et al [53] has proposed a novel method for oblivious and robust image watermarking scheme using Multiple Descriptions Coding (MDC) and Quantization Index Modulation (QIM).Proposed scheme is characterized with Blocked Discrete Hadamard Transform (DHT) parameters and Genetic Algorithm (GA) is used for parameter optimization. The performance improvement is achieved over the existing methods in terms of Peak Signal to Noise Ratio (PSNR) and Normalized Cross correlation (NCC). The tradeoff between the transparency and robustness is considered as an optimization problem and is solved by applying

Genetic Algorithm proposed by P.Surekha et al [54]. The amplification factor and robustness against attacks. For the Colour image , the image is decomposed into their colour components viz, R, G and B. One of these matrices are divided into odd and even banks of 8X8 each, such that pixel values present in the odd positions will go to the odd bank and similarly even bank. The data to be embedded is embedded into one of the banks say odd bank. The embedding position of the data to be watermarked into the image is found out using Genetic Algorithm given by Abduljabbar Shaamala et al [55] and in [57] he gives the effect of embedding data in frequency domain on the robustness in genetic watermarking.

D. Venkatesan et al [56] has used a center of mass selection operator based Genetic Algorithm, to investigate the variation of maximum fitness based on the higher PSNR value of watermarked image, against embedding strength, number of genes, various payload of digital watermark. Azman Yasin et al [58] has given a study of comparison between two existing methods: Dual Intermediate Significant Bit (DISB) and Genetic Algorithm (GA). GA is used in determining the minimum fitness value in which the fittest is the absolute value between the pixel and chromosome. It produces a high quality watermarked image, but there is a big difference in the processing time, so the DISB method is faster than the GA method. In particular, IR is one of the key steps in medical imaging, with applications ranging from computer assisted diagnosis to computer aided therapy and surgery. IR can be formulated as an optimization problem on a real coded genetic algorithm with a more appropriate design given by Andrea Valsecchi [59].

5. ATTACKS AND EMBEDDING CAPACITY

When the images are transmitted there is the possibility of attack either intentionally or unintentionally. This may degrade the quality of the image and affect the performance of the system. Hence the watermarking system should be robust enough to survive the attacks. In order to evaluate the robustness and effectiveness of our watermarking method, it is necessary to investigate the influence of different attacks on image as innocent Attacks and Malicious attacks [60]. A watermarking technique in wavelet domain for the EHR data is based on energy band selection [61] [62] and in reference to the bit location in the reference image. Virtually all lossless embedding techniques increase the file size of the embedded image. The lossless embedding methods that preserve the file size of images in the RLE encoded BMP and the JPEG formats. A reversible method can be used to embed information without increasing it size.

6. PROSPECTS & APPLICATIONS

Digital image watermarking has numerous applications in variety of fields such as Copyright Protection, Content Archiving, Meta data Insertion, Tamper Detection, Digital Fingerprinting. Depending on the medical application area (health, administrative, teaching, research ...) the trade-offs among robustness, invisibility and capacity varies. The image segmentation is applied to select only the required portion of medical image. RONI approaches will leave intact the diagnostic information, but they can be applied only if a RONI exists. Furthermore, the capacity is dependent upon the RONI area size. Neural networks can be used to implement an automated system of creating maximum-strength watermarks. Artificial intelligence techniques can be used for watermarking to have better robustness against many image processing attacks like rotation, sharpening, image contrast attacks. Fuzzy-Neuro system to offer combined advantages of both Artificial Neural Networks and Fuzzy Logic.

7. CONCLUSION

A review of various watermarking embedding techniques have been done which suits for medical images with higher embedding capacity and higher imperceptibility with which it can withstand against attacks as well. Since no compromise can be made on the fidelity criteria of the medical images appropriate transforms for medical image could be identified and incorporated to bring an optimal embedding of diagnosed data in medical images. Analysis of various detection algorithm also has been done for medical images which was optimized using suitable genetic algorithm. There was always been a tradeoff between robustness, capacity and imperceptibility hence the embedding and detection algorithm has to suit for the need of data retrieval and archiving.

ACKNOWLEDGEMENT

I would like to thank Dr. R.Varatharajan, Research Guide for his continuous support and ideas for my research work. I would also like to thank Dr.J.Samuel Manoharan, for his complete guidance regard the work .

REFERENCES

[1] Asabe, S. A., Oye, N. D., Monday Goji," Hospital Patient Database Management System A Case Study Of General Hospital North-Bank Makurdi- Nigeria" , Compusoft, An International Journal Of Advanced Computer Technology, 2 (3), Volume-II, Issue-III, pp-65-72, 2013

[2] G. Coatrieux, L. Lecornu, B. Sankur, Members, IEEE, Ch. Roux, Fellow, IEEE ," A Review Of Image Watermarking Applications In Healthcare" , A Proceedings Of The 28th IEEE EMBS Annual International Conference New York City, pp-4691-4694, 2006

[3] Adol Esquivel, Dean F Sittig, Daniel R Murphy And Hardeep Singh ," Improving The Effectiveness Of Electronic Health Record-Based Referral Processes" , pp-1-8, 2012 .

[4] Prashant Vaidyanathan – Nitish Malhotra – Jagadish Nayak , "A New Encryption Technique For The Secured Transmission And Storage Of Text Information With Medical Images", Engineering Review Vol. 32, Issue 1, pp-57-63, 2012

[5] Cherif Moumen, Malek Benslama, and Mekhilef Saad, "Cryptography of the Medical Images", PIERS Proceedings, pp-42-48, 2012

[6] Lahieb Mohammed Jawad and Ghazali Bin Sulong , " A Review Of Color Image Encryption Techniques" , IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 6, No 1, pp-266-275,2013

[7] Quist-Aphetsi Kester, MIEEE ," A Visual Cryptographic Technique For Securing Medical Images" , International Journal of Emerging Technology and Advanced Engineering Volume 3, Issue 6, pp-496-500,(2013)

[8] Rohaya Mohd-Nor , " Medical Imaging Trends And Implementation: Issues And Challenges For Developing Countries ",Journal Of Health Informatics In Developing Countries, pp-89-98, 2011

[9] Li-Chin Huang, Lin-Yu Tseng, And Min-Shiang Hwang , " The Study On Data Hiding In Medical Images " , International Journal Of Network Security, Vol.14, No.6, Pp.301-309 , 2012

[10] Mohamed Ali Hajjaji, El-Bay Bourennane, Abdellatif Mtibaa, Gilberto Ochoa-Ruiz , " A Digital Watermarking Algorithm Based On Quantization Of The DCT: Application On Medical Imaging" , pp-372-377, IEEE ,2013

[11] Arathi Chitla And Chandra Mohan M , " Authenticating Medical Images With Lossless Digital Watermarking" , International Journal Of Multidisciplinary And Current Research , Vol 2, pp-291-296,2014

[12] Mohamed M. Abd-Eldayem , "A proposed security technique based on watermarking and encryption for digital imaging and communications in medicine " , Egyptian Informatics Journal , pp- 1-13,2013

[13] Neha S. Korde, Dr. A. A. Gurjar , " Wavelet Based Medical Image Compression For Telemedicine Application" , American Journal of Engineering Research (AJER) Volume-03, pp-106-111,2014
[14] Mohamed Ali, Hajjaji Abdellatif Mtibaa, El-Bey Bourennane , "A Watermarking of Medical Image: Method Based "LSB" " , Journal of Emerging Trends in Computing and Information Sciences , VOL. 2, NO. 12, pp-714-721,2011
[15] A. Siva Sankar , T. Jayachandra Prasad , M.N. Giri Prasad, "LSB Based Lossless Digital Image Watermarking using Polynomials in Spatial Domain for DRM " , 2nd International Conference and Workshop on Emerging Trends in Technology (ICWET), pp-18-24, 2011
[16] Mohamed Radouane, Tarik Boujiha, Rochdi Messoussi, Nadia Idrissi, Ahmed Roukh, "A Method Of LSB Substitution Based On Image Blocks And Maximum Entropy" , IJCSI International Journal Of Computer Science Issues, Vol. 10, Issue 1, No 1, pp-371-374, 2013
[17] Hirak Kumar Maity, Santi Prasad Maity, "Joint Robust and Reversible Watermarking for Medical Images" , 2nd International Conference on Communication, Computing & Security [ICCCS], Vol 6,pp-275-282,2012
[18] Mr. Venugopal Reddy .Ch, Dr. Siddaiah.P , "Hybrid LWT- SVD Watermarking Optimized Using Metaheuristic Algorithms Along With Encryption For Medical Image Security " , Signal & Image Processing : An International Journal (SIPIJ) Vol.6, No.1, pp-75-95, 2015
[19] S. A. Hosseini and S. Ghofrani , "Using Contourlet Transform and Discrete Cosine Transform and SVD for Digital Watermarking " , ACSIJ Advances in Computer Science: an International Journal, Vol.2, Issue 5, No.6 ,pp-20-28, 2013
[20] Rajendra Acharya U.A,U.C. Nir anjanb, S.S. Iyengarc, N. Kannathala, Lim Choo Mina, "Simultaneous Storage Of Patient Information With Medical Images In The Frequency Domain", Elseiver, Computer Methods And Programs In Biomedicine 76, pp-13—19, 2004
[21] Federica Battisti, Michela Cancellaro, Giulia Boato, Marco Carli,And Alessandro Neri (Eurasipmember),"Jointwatermarking And Encryption Of Color Images In The Fibonacci-Haar Domain ",Hindawi Publishing Corporation Eurasip Journal On Advances In Signal Processing , pp-1-9,Volume 2009.
[22] Gouenou Coatrieux, Member, IEEE, Clara Le Guillou, Jean-Michel Cauvin, And Christian Roux, Fellow, IEEE, "Reversible Watermarking For Knowledge Digest Embedding And Reliability Control In Medical Images" , IEEE Transactions On Information Technology In Biomedicine, Vol. 13, No. 2,pp-158-165,2009
[23] Sofiane Braci, R´Emy Boyer And Claude Delpha ,"Security Evaluation Of Informed Watermarking Schemes ", pp-117-120,2009 IEEE, ICIP
[24] Wei Pan, Gouenou Coatrieux, Nora Cuppens-Boulahia, Fr´Ed´Eric Cuppens And Christian Roux, "Watermarking To Enforce Medical Image Access And Usage Control Policy", pp- 251-260, 2010 IEEE, IEEE Computer Society.
[25] Mohammad-Saleh Nambakhsha, Alireza Ahmadianb, Habib Zaidic, "A Contextual Based Double Watermarking Of Pet Images By Patient ID And ECG Signal, Computer Methods And Programs In BioMedicine Vol 4, pp- 418–425 , 2 0 1 1
[26] Xiao-Wei Li, Sung-Jin Cho, And Seok-Tae Kim," 2-D CAT-Based Medical Image Watermarking Algorithm", International Journal Of Computer Theory And Engineering, Vol. 4, No. 5,pp- 722-725, 2012
[27] Dalel Bouslimia,B, Gouenou Coatrieuxa,B,Christian Rouxa,B , "A Joint Encryption/Watermarking Algorithm For Verifying The Reliability Of Medical Images: Application To Echographic Images ",Computer Methods And Programs In BioMedicine ,Vol 06 , pp-47–54, 2 0 1 2
[28] Dalel Bouslimi, Member, IEEE, Gouenou Coatrieux, Member, IEEE, Michel Cozic, And Christian Roux, Fellow, IEEE," A Joint Encryption/Watermarking System For Verifying The Reliability Of Medical Images", IEEE Transactions On Information Technology In Biomedicine, Vol. 16, No. 5,pp-891-899, 2012
[29] M.V.S.S.Babu , "A Robust Watermarking Algorithm For Image Authentication" , 2012 International Conference On Information And Network Technology (ICINT 2012) , Vol. 37, pp- 220-227, 2012

[30] Vinay Pandey, Manish Shrivastava , " Secure Medical Image Transmission Using Combined Approach of Data-Hiding, Encryption And Steganography", International Journal Of Advanced Research In Computer Science And Software Engineering , Volume 2, Issue 12,pp-54-57 , 2012

[31] Mohamed M. Abd-Eldayem ," A Proposed Security Technique Based On Watermarking And Encryption For Digital Imaging And Communications In Medicine" , Egyptian Informatics Journal, Vol 14, pp- 1-13 , (2013)

[32] Hung-I Hsiao , Junghsi Lee ," Cryptographic System For Medical Image Security Using Chaotic APFM Nonlinear Adaptive Filter ", pp-131- 134 , 2013 IEEE

[33] Lamri Laouamer , Laurent T. Nana , Muath Al Shaikh , Anca C. Pascu , " Informed Symmetric Encryption Algorithm For Dicom Medical Image Based On N-Grams" , Science And Information Conference ,pp-353-357, 2013

[34] A. Kannammal, S. Subha Rani , "Two Level Security For Medical Images Using Watermarking/Encryption Algorithms" ,Vol 24, pp- 111-120, 2014 Wiley Periodicals, Inc.

[35] P. Viswanathan, Member, IEEE, And P. Venkata Krishna, Senior Member, IEEE ," A Joint FED Watermarking System Using Spatial Fusion For Verifying The Security Issues Of Teleradiology" , IEEE Journal Of Biomedical And Health Informatics, Vol. 18, No. 3,pp- 753-764, 2014

[36] Shun Zhang, Tiegang Gao, And Lin Gao ," A Novel Encryption Frame For Medical Image With Watermark Based On Hyperchaotic System", Hindawi Publishing Corporation,Mathematical Problems in Engineering , Volume 2014, pp- 1- 11 , 2014

[37] V.Amutha, C.T. Vijay Nagaraj ," A Secured Joint Encrypted Watermarking In Medical Image Using Block Cipher Algorithm" , International Journal Of Innovative Research In Science, Engineering And Technology Volume 3, Special Issue 3,pp-1099-1104, 2014

[38] Wei-Hung Lin A, Yuh-Rau Wang B, Shi-Jinn Horng A,C, Tzong-Wann Kao D, " A Blind Watermarking Method Using Maximum Wavelet Coefficient Quantization" , Yi Pan E , Expert Systems With Applications Vol -36, pp- 11509–11516,2009

[39] Xinge You, Senior Member, IEEE, Liang Du, Member, IEEE, Yiu-Ming Cheung, Senior Member, IEEE, And Qiuhui Chen , "A Blindwatermarking Scheme Using New Nontensor Product Wavelet Filter Banks" , IEEE Transactions On Image Processing, Vol. 19, No. 12, 2010

[40] Deepa Mathew K, "Evolutionary Computation For Optimization Techniques SVD Based Image Watermarking Scheme ", IJCA, 2010

[41] Swanirbhar Majumder , Monjul Saikia , Tirtha Sankar Das, Subir Kumar Sarkar, "Hybrid Image Watermarking Scheme Using SVD Based Contourlet Transform", International Conference On Computer & Communication Technology (ICCCT)-2011

[42] Jila Ayubi, Shahram Mohanna, Farahnaz Mohanna And Mehdi Rezaei ," A Chaos Based Blind Digital Image Watermarking In The Wavelet Transform Domain " , IJCSI International Journal Of Computer Science Issues, Vol. 8, Issue 4, No 2, 2011

[43] Surya Pratap Sing, Paresh Rawat, Sudhir Agrawal, "A Robust Watermarking Approach using DCT-DWT", International Journal Of Emerging Technology And Advanced Engineering ,Volume 2, Issue 8, 2012

[44] Yinglan Fang, Lin Tian , "An Improved Blind Watermarking Algorithm For Image Based On DWT Domain " , Journal Of Theoretical And Applied Information Technology ,Vol 45, pp- 168-173, 2012.

[45] Farhad Saeed, Zahedan, Mehdi Golestanian, Mohamadreza Azimi, " A Blind Watermarking Algorithm Based On DCT-DWT And Arnold Transform",Department Of Telecommunications, IJCSE, Vol. 2 , No.06, pp- 328-334, 2013

[46] Nidhi Bisla, Prachi Chaudhary , "Comparative Study Of DWT And DWT-SVD Image Watermarking Techniques " , International Journal Of Advanced Research In Computer Science And Software Engineering , Volume 3, Issue 6,pp-821-825, 2013

[47] Sudeb Das, Malay Kumar Kund , "Effective Management Of Medical Informationthrough ROI-Lossless Fragile Image Watermarking Technique" , C omputer Methods And Programs in BioMedicine , pp-1-14, 2013-Elsevier

[48] J.B. Lima A, E.A.O.Lima B, F.Madeiro B ,"Image Encryption Based On The Finite Field Cosine Transform", 2013, Signal Processing: Image Communication , Elsevier, pp- 1-14

[49] N.Venkatram, L.S.S.Reddy, P.V.V.Kishore, " Blind Medical Image Watermarking With LWT – SVD For Telemedicine Applications" , WSEAS Transactions On Signal Processing, Vol 10,pp-288-300, 2014

[50] B.Jagadeesh, D.Praveen Kumar, "Robust Digital Image Watermarking Scheme Based On DCT And BPNN " , International Journal Of Advanced Research In Electrical Electronics And Instrumentation Engineering , Vol. 3, Issue 5,pp-9453-9459, 2014

[51] Hamed Modaghegh, Hossein Khosravi R., Mohammad- R. Akbarzadeh-T ," A New Adjustable Blind Watermarking Based On GA And SVD" , Innovations'09: 6th International Conference On Innovations in Information Technology, pp-6-10, 2009

[52] Veysel Aslantas A, Saban Ozer B, Serkan Ozturk A," Improving The Performance Of DCT-Based Fragile Watermarking Using Intelligent Optimization Algorithms" , Optics Communications,Vol 282,pp-2806-2817, (2009), Elsevier

[53] K. Ramanjaneyulu, K. Rajarajeswari ," An Oblivious And Robust Multiple Image Watermarking Scheme Using Genetic Algorithm", The International Journal Of Multimedia & Its Applications (IJMA) Vol.2, No.3,pp- 19-38 , 2010

[54] P. Surekha And S. Sumathi, " Implementation Of Genetic Algorithm For A DWT Based Image Watermarking Scheme", Journal On Soft Computing: Special Issue On Fuzzy In Industrial And Process Automation, Volume: 02, Issue: 01, pp – 244- 252, 2011

[55] Abduljabbar Shaamala, Shahidan M. Abdullah, And Azizah A. Manaf ," The Effect Of DCT And DWT Domains On The Robustness Of Genetic Watermarking ", : ICIEIS 2011, Part I, CCIS 251, Pp. 310–318, 2011. Springer 2011

[56] D. Venkatesan , K. Kannan And S. Raja Balachandar , "Optimization Of Fidelity In Digital Image Watermarking Using A New Genetic Algorithm" , Applied Mathematical Sciences, Vol. 6, 2012, No. 73,pp- 3607 - 3614

[57] Abduljabbar Shaamala, Azizah A. Manaf , "Study Of The Effected Genetic Watermarking Robustness Under DCT And DWT Domains", International Journal On New Computer Architectures And Their Applications (IJNCAA) 2(2): pp-353-360 , The Society Of Digital Information And Wireless Communications, 2012

[58] Azman Yasin, Akram M. Zeki, And Ghassan N. Mohammed ," A Comparison Of Watermarking Image Quality Based On Dual Intermediate Significant Bit With Genetic Algorithm ", Proceedings Of The 4th International Conference On Computing And Informatics, ICOCI 2013 Paper No.008

[59] Andrea Valsecchi And Sergio Damas , Jos´E Santamar´Ia , Linda Marrakchi-Kacem ," Genetic Algorithms For Voxel-Based Medical Image Registration", 2013 Fourth International Workshop On Computational Intelligence In Medical Imaging (CIMI)

[60] J.-B. Aupet, E. Garcia, H. Guyennet, J.-C. Lapayre, and D. Martins , "Security in Medical Telediagnosis" , Springer 2010

[61] Dr. Rajendra D. Kanphade and N.S. Narawade, " Forward Modified Histogram Shifting based Reversible Watermarking with Reduced Pixel Shifting and High Embedding Capacity", . International Journal of Electronics and Communication Engineering. ISSN 0974-2166 Volume 5, Number 2 (2012), pp. 185-191

[62] K.Anusudha , N.Venkateswaren, "Energy Based Wavelet Domain Medical Image Watermarking" , International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering Vol. 3, Issue 2,pp-7132-7140, 2014

[63] Sameh Oueslati, Adnene Cherif , Bassel Solaimane , " Adaptative Image Watermarking Scheme Based On Neural Network", International Journal Of Engineering Science And Technology (IJEST) , Vol. 3, No. 1,pp-748-756, 2011

AUTHORS

Mrs.S.Priya, research scholar in Bharath University, Chennai.Her area of research is in Medical Image Processing, Electronic patient record, and Medical image watermarking.

Dr.R.Varatharajan,Professor &Head of ECE,Sri Lakshmi Ammal Engineering College. He has completed his PhD from Bharath University ,Chennai.He has published more journals in International and National Level.His research areas are VLSI,Medical Image Processing.

OPTIMIZATION OF AVERAGE DISTANCE BASED SELF-RELOCATION ALGORITHM USING AUGMENTED LAGRANGIAN METHOD

Shivani Dadwal[1] and T. S. Panag[2]

[1]Deptt of Electronics and Communication Engineering, BBSBEC,
Fatehgarh Sahib, India
shivanidadwal@gmail.com
[2]Deptt of Electronics and Communication Engineering, BBSBEC,
Fatehgarh Sahib, India
tripatjot.singh@bbsbec.ac.in

ABSTRACT

Mobile robots with sensors installed on them are used in wireless sensor networks to generate information about the area. These mobile robotic sensors have to relocate themselves after initial location in the field to gain maximum coverage The average distance based algorithm for relocation process of mobile sensors does not require any GPS system for tracking the robotic sensors, thus avoiding cost, but increasing energy consumption. Augmented Lagrangian method is introduced in average distance based algorithm to reduce the extra energy consumption by sensors in average distance based relocation process. This modified average distance relocation scheme also improves the coverage area and the time taken by mobile robotic sensors to come to their final positions.

KEYWORDS

Average distance based self-relocation, modified average distance self-relocation.

1. INTRODUCTION

A lot of work has been done on sensor network as it has a wide application in various fields. One of its major application is in military area wherein surveillance of enemy areas has to be done. Earlier static sensors were deployed for detecting targets in a particular area. This reduced the coverage area of sensors. With advances in mobile robotics, now sensors can be carried by robotic structures. All they need is a direction which is administered by an algorithm. Execution of these algorithm leads to deployment of sensors in an area in a manner that maximum area is covered and all the targets are well detected. Even civilian application of such mobile sensors exists in property and homeland security [1, 2].

David C. Wyld et al. (Eds) : ACITY, DPPR, VLSI, WiMNET, AIAA, CNDC - 2015
pp. 193–203, 2015. © CS & IT-CSCP 2015 DOI : 10.5121/csit.2015.51317

For military purposes, the sensors are mostly dropped from air, which leads to a random deployment of sensors. These sensors have to reposition themselves. This issue has been covered in our paper. Optimization of coverage has been an active topic in sensor networks as along with coverage there are many other issues like minimum energy consumption and maximum lifetime which have to be kept in mind. Optimization of static sensor networks has been done using Genetic Algorithms which helps in locating sensors to their best position for maximum coverage and saves energy resulting in increased lifetime [3,4,5,6]. Other than this a few distributed and centralized algorithms have been introduced in an effort to modify and improve sensor networking [7]. The Genetic algorithm for optimization cannot be used for distributed algorithms as they require a central operating system to control the sensor positioning. Moreover Genetic algorithm can be used only where the area is well known.

Many algorithms have been developed for placing the sensors evenly in a field, out of which potential field algorithm is one [8]. In this algorithm, a potential field is generated by the sensors and they move accordingly. Another algorithm is the virtual force algorithm [9] in which the sensors move as per the attractive and repulsive forces generated by the sensors depending on the distances between them and attain their final positions. Some other methods namely, density control method [10] and fluid model based method [11] have also been introduced. There is another algorithm that repairs the coverage by finding the terminated sensors and moves these sensors to uncovered areas [12]. These algorithms listed above are applicable for distributed network where there is no central node and each sensor decides its own path. For this kind of network, a GPS system is must in order to track down the position of sensors after and before relocation. GPS is global positioning system that consumes a lot of power and is costly. Every time the sensor move during the execution of their respective algorithms, they send information to the GPS system which again consumes power.

Some methods were introduced to decrease the power consumption of sensor out of which one was to reduce the sensing range of sensor [13]. This was done on non-mobile sensors. It saves energy and increases the lifetime of sensors. The average distance based relocation process does not use any GPS system and hence cuts the energy consumption and cost as well [14]. But since there is no GPS used, the sensors consume extra energy to find their best final position which also increases the number of iterations to find the final coverage. Hence, optimization of average distance based self-relocation process is introduced to facilitate the sensors to find their final positions in less time which would also reduce energy consumption. This is done using augmented lagrangian optimization method [19]. In the next section, average distance based self-relocation algorithm has been discussed.

2. AVERAGE DISTANCE BASED SELF RELOCATION PROCEESS

2.1. Assumptions

Following are a few assumptions that we consider for the algorithm:

- Each sensor has a sensing area in the form of a circle, having radius, r. The probability of covering this area is 1.
- The area A, where the sensors are randomly deployed is known by the sensors approximately

- If the sensors come within their sensing range Rc, the strength of the signals transmitted by each sensor can be measured by the other.
- All sensors have certain range of communication and have a transmission power.
- Sensors have the ability to move as per the coordinates given after execution of algorithm.
- Sensors can detect obstacles in the field.
- Sensors that meet an obstacle in the field has its movement blocked and cannot communicate with the rest of sensors.

2.2. Framework

The main aim of this algorithm is designing a distributed algorithm which has a self- relocation capability to optimize the coverage area of field using less energy. It this algorithm, the distances between the sensors have to be known in order to relocate the sensors. The sensors transmit signal in the field once they are randomly deployed. These signals are intercepted by the sensors that come within the reach of .sensor areas of other sensors. The received signal strength is measured and corresponding distance is known.

Firstly, a "hello" signal is transmitted by the sensors which gives the signal strength to all the sensors, near or far, lying in sensor range. The distance corresponding to this signal strength is calculated by the sensors. Taking these distance information into account, the sensors move towards or away to each other and relocate themselves. Also, any obstacle coming in way has to be avoided by the sensors.

3. METHODOLOGY

3.1. Ideal Deployment

The ideal deployment is achieved when there are no spaces between the sensors. Such a condition is possible when the distance between the sensors is $\sqrt{3}r$ [15]. This is shown in Figure 1. Since it is not possible to achieve an ideal condition, in this algorithm, we try to achieve a near to ideal condition by placing sensors close. But as the field may have a deformity of obstacles, and various other factors, such a condition is difficult to achieve.

3.2. Calculation of Threshold

The total number of sensors in the field and the sensing field area are used to calculate the threshold distance dth and sensing radius that are near to ideal deployment. The threshold distance dth decides the sensor movement.

Let the total area of the sensor field be A. As shown in Figure 1, assume that each sensor has an effective area of coverage, E as in [14]. Let the total number of sensors deployed in the field be N. Effective coverage of each sensor is given as:

$$E = \frac{A}{N} \qquad (1)$$

Also, E is a hexagonal area. Area of a hexagon is given as:

$$E = \frac{3\sqrt{3}}{2}.r^2 \quad (2)$$

The threshold distance and sensing radius are calculated by the following equations:

$$r = \sqrt{\frac{2}{3\sqrt{3}}.\frac{A}{N}} \quad (3)$$

$$d_{th} = \sqrt{3}r \quad (4)$$

The effective coverage, E should be larger than this value, as the sensors which lie close to obstacles or to the edges will have less coverage. Hence, the threshold distance and sensor radius are increased by 15%.

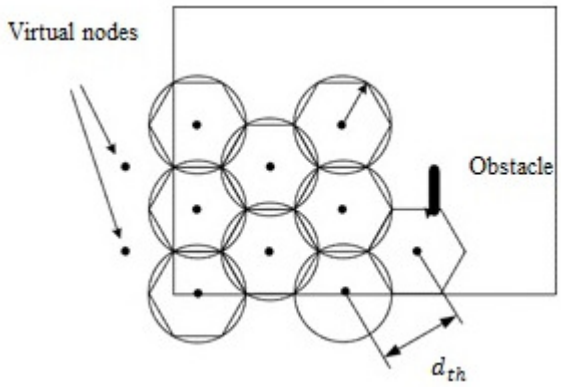


Figure 1. : Ideal coverage

3.3. Virtual Nodes

The algorithm considers that there exist virtual nodes at the boundary of the field. This is done to make deployment easy for the sensors lying close to the edges. Such virtual nodes do not actually exist. They are just used to avoid the sensors from getting closer to the edges. As shown in Figure 1, virtual nodes are considered along the boundary. No virtual nodes are required in the optimization technique used in this work.

3.4. Movement Standards

The sensors relocate themselves by adjusting the distances between themselves. They either move far or get closer to each other. No sensor has information about the direction of the other sensor. The criteria in which the sensors move is described as below:

Standard 1: If there is at least one sensor in the communication range of sensor S having distance less than $0.9d_{th}$, then the sensor will move away from other sensors.

Standard 2: If the standard 1 is not met and not more than 2 sensors lie at distance less than $1.1d_{th}$ from S, then the sensor needs to move closer to other sensors.

Standard 3: Sensor S need not move if the above two standards are not met.

A 10% margin is kept along these standards, so that the sensor is able to achieve a distance nearer to d_{th} from the rest of sensors.

3.5. Moving Distance

The movement of sensors is based on two standards, hence the moving distance is calculated by the following equation:

$$d_{travel} = \begin{cases} d_{th} - \frac{1}{m_1}\sum_{j=1}^{m_1} d_j & for\ standard\ 1 \\ \frac{1}{m_2}\sum_{i=1}^{m_2} d_i - d_{th} & for\ standard\ 2 \\ 0 & for\ standard\ 3 \end{cases} \quad (5)$$

In the above equation, d_j and d_i are the distances from sensor S to other sensors. In the first standard, the sensor moves only for the sensors closer than distance threshold. The total number of sensors is given by m_1. In the second standard, all the neighboring sensors to sensor S are taken into account. The total number of sensors is m_2.

The direction of movement of the sensor is chosen randomly as the sensor is unaware of the neighbouring sensors. There can be a back and forth movement of sensors. To avoid this direction control scheme is used. As the sensors move, the difference of direction of movement is kept less than 90 degrees. Let the last direction of movement be α, then in the next movement direction has to be in between α-90 degrees to α+90 degrees. When standard 1 is executed, the sensors move away from each other and when standard 2 is executed, the sensors move nearer to each other. As the sensors are moving to the direction chosen randomly, they check after moving through a short distance if the required coverage is attained. If not, they come back to their original positions.

4. SIMULATION AND ANALYSIS

For examining our results, we consider that the sensing field is a 100 by 100 grid structure. Each grid is 1 meter apart from the other grid. Consider that the sensing range of the sensors used lies within 18 to 25 meters. Hence, the maximum sensing radius is 25 meters. The range of communication of sensors is kept almost double the maximum sensing radius i.e. 55 meters.

4.1. A Average Distance Based Self- relocation Algorithm Performance

Let 20 sensors be randomly deployed in the sensor field. The distance threshold and sensing radius can be calculated from equations (3) and (4). A 15% increase should also be considered as explained in Section 2.3.

The following equations give the sensing radius and threshold distance of 20 sensors:

$$r = 1.15\sqrt{\frac{2}{3\sqrt{3}}.\frac{A}{N}} = \sqrt{\frac{2}{3\sqrt{3}}\frac{100\times100}{20}} = 15.9\ meters$$

$$d_{th} = \sqrt{3}r = \sqrt{3} \times 15.9 = 27.6\ meters$$

To analyze the results, three different conditions of initial sensor placement can be considered. In the first case, the sensors are all placed in the center of the sensing field such that they cover 50 by 50 meters area. In the second condition, the sensors are deployed or scattered in the whole sensing area. In the third condition, sensor are divided into 2 groups which are separately placed in the sensing field. The coverage initially is calculated for all the conditions which come out to be 55%, 50% and 61% approximately. The coverage can be calculated by the following equations:

$$R_{coverage} = \frac{A_{covered}}{A_{total}} \quad (6)$$

Here,

$R_{coverage} = coverage\ ratio$

$A_{covered} = area\ covered\ by\ sensors\ cooperatively$

$A_{total} = sensing\ field\ area$

If we have to find out the coverage from the 100 by 100 grid structure, then the following equation can be used:

$$R_{coverage} = \frac{n}{N} \quad (7)$$

Here,

n= number of grid points covered by sensors

N= total number of grid points.

4.2. Coverage Analysis

The sensors in this algorithm do not have a fixed direction of movement. They move randomly and check if the required conditions are met. So, each of the three conditions stated in the above section are run 10000 times with a desired round number of 20. The execution results are shown in the Figure 2.

In figure 2, the results of average distance based algorithm are compared with virtual force algorithm. The virtual force algorithm is executed by taking its parameters into account as given in [16]. In all the three conditions, the coverage increases as the number of rounds increase. An average distance based self-relocation algorithm can achieve nearly 94% of coverage in the sensing field after 20 rounds. For a virtual force algorithm, after 20 rounds 93% coverage is achieved.

Thus, both the algorithm lead to almost same coverage of sensing field by the sensors. The major difference is that the average distance based algorithm does not require a GPS hardware, reducing the cost of the sensor network. It can hence be used in those areas where GPS cannot operate, like in under water systems.

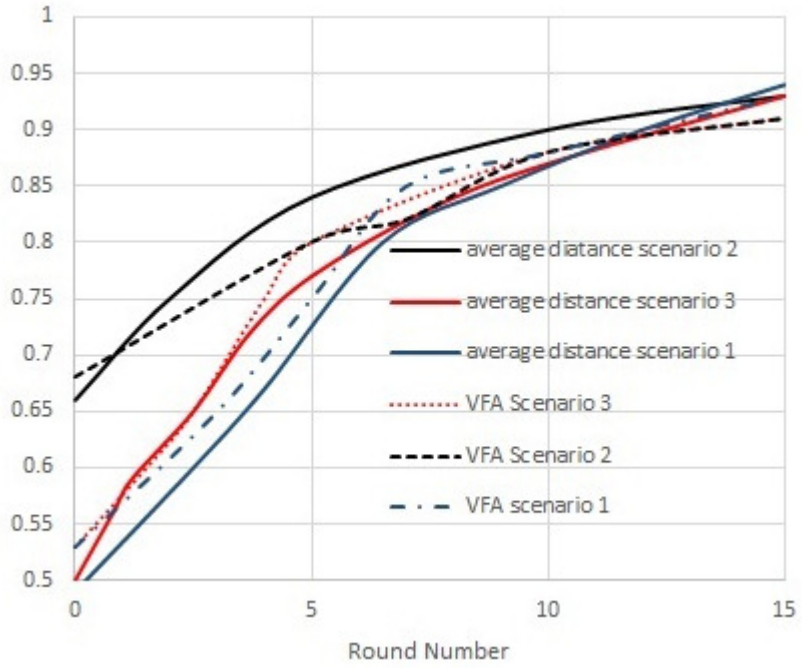


Figure 2: Simulation results of average coverage vs round number for average distance based self – relocation algorithm

4.2. Energy Analysis

An energy analysis model has been discussed in [17]. This model has been used for mobile robots as per which energy consumed by a robot to move 1 meter is equal to 9.34 Joules if it is moving at a speed of 0.08 m/s constantly. The amount of energy used by robots to turn by 90 degrees is 2.35 Joules. Both travelling and turning of mobile robots is considered to be at a constant speed. By doing so, we can plot a linear graph between energy consumed by mobile robot and coverage. As per the above discussion, we can divide energy consumption into two following parts:

- *Energy used while travelling:* If the mobile robot keeps moving at a constant rate in one direction, then the energy consumed for one single sensor node in a single round is given as:

$$E_{travel} = d_{travel} \times 9.34(Joules)$$

- *Energy used in direction changing:* As discussed in the relocation algorithm, the sensor move randomly in some direction and checks its position by recalculating the signal strength. After this, it decides whether to move back to original position or not. Hence, during this process, the energy used by the sensor to turn in some direction is given as:

$$E_{turn} = \begin{cases} (A_{diff}/90) \times 2.35(Joules) & Keep\ moivng \\ (360/90) \times 2.35(Joules) & Turning\ back \end{cases}$$

In the above equation, A_{diff} is the difference of direction between the previous and later direction of sensor movement. If the new position does not satisfy the coverage requirement of sensor, then it gets back to position where it started by moving at 360 degrees, as shown in the above equation.

To plot a graph, virtual force algorithm has also been considered. Its energy consumption is calculated. In the virtual force algorithm, the above two energy consumption are added to the energy used by the GPS system in locating the sensor nodes and in exchange of information between sensor and GPS. The GPS chip uses 198 MW as in [18]. As the sensors are moving constantly at 0.08m/s, the GPS consumes energy per meter given by following equation:

$$0.198 \times (1/0.08) = 2.475\, Joules/meter$$

As given in the Figure 3, virtual force algorithm will consume lesser energy as compared to average distance based algorithm. The graph is plotted between average energy consumption and average coverage for both the algorithms. The VFA uses lesser energy as it is GPS enabled, making it easier for the sensor to redeploy themselves faster than the sensors in average distance based algorithm. In VFA sensors take less time to redeploy as they know their positions corresponding to other sensors and also know the locations of rest of the sensors in the field.

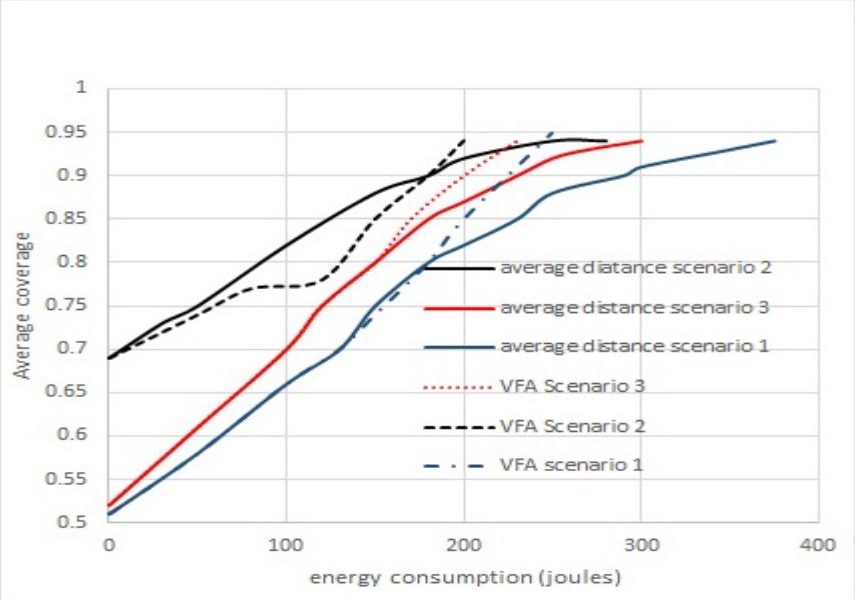


Figure 3: Simulation results for average coverage vs energy consumption

5. OPTIMIZATION USING AUGMENTED LANGRANGIAN METHOD

The basic drawback of average distance based algorithm is that it consumes more energy in relocating the sensors as they move back and forth many times to come to an appropriate position to get a good coverage.

Here, if an optimization technique is used to help sensors in their relocation process, they can come to final position in lesser time consuming lesser energy. In the average distance based self-relocation process, virtual nodes are considered at the boundary and outside the boundary as shown in figure 1. Using these virtual nodes, the sensors make an idea of their boundary and the area beyond which they are restricted. The sensors move back and forth in self –relocation process, during which the sensors nearer to the boundary have to recalculate its positions to stay

within the boundary, which can be energy consuming. Thus, in the augmented lagrangian method, the sensors are applied a penalty function. Using this penalty function, if the sensor skips outside the boundary by certain distance d_{out}, it is made to to come inside the boundary by the same distance d_{out}, as measured from the boundary. During this process, the sensor might overlap or come into the boundary of another sensor, violating the threshold distance conditions as discussed in section 2.3.

In the augmented lagrangian optimization method, an objective function has to be considered. The minimum distance between the sensors should be d_{th} , as in average distance based relocation process. This minimum distance might change when penalty function is applied, so minimum distance between sensor i.e. threshold distance d_{th} is the objective function subject to constraints in augmented lagrangian optimization method.

Now let the boundary of the sensor field be defined by following functions.

$$g_1 > 0, g_2 \geq 0, g_3 < 0, \dots, g_n > 0$$

Here, $g_1, g_2, g_3, \dots, g_n$ are the boundary equations for sensor field and n is the number of boundaries.

In augmented lagrangian optimization method, new objective function is given as

$$f[X] - \alpha\langle d_{th}\rangle - \beta\langle g_1[X] + g_2[X] + \cdots + g_n[X]\rangle$$

$[X]$ is the co-ordinate of all the sensors obtained after first iteration of average distance based relocation algorithm. The outline of the algorithm is as follows:

- The $[X]$ co-ordinates of all the sensors in the sensing field are obtained after the execution of first iteration of average distance based relocation algorithm.
- The boundary constraint violations are checked i.e. it is checked if any sensor is lying on the boundary or beyond the boundary. The constraints are checked n times which is equal to the number of sensors in the field.
- If any constraint is violated, let us say for nth sensor,$[g_1]_i \not< 0$, which means that constraint g_1 is violated by ith sensor and the sensor is moving beyond the boundary of sensing field. Hence, the distance D_i is calculated between g_1 and $[X, Y]_i$. X and Y are the co-ordinated of ith sensor lying outside the boundary.
- To compensate for distance and to bring the sensor back inside the field, negative of distance D_i is added in g_1 and the new co-ordinates of ith sensor are calculated.
- Now, all the co-ordinates serve as initial solution for average distance based algorithm which is run again.
- The threshold distance are again checked in the self -relocation algorithm.

Hence, this method is a modified version of average distance based self-relocation algorithm using the properties of augmented lagrangian optimization technique.

6. FUTURE WORK

The biggest advantage of the relocation scheme discussed is non- requirement of a GPS hardware which cuts cost and increases the applicability of this relocation process where GPS cannot be used. The drawback of more energy consumption is to some extent improved by the optimization technique, yet more work can be done to further lessen the energy consumption of non GPS using sensors.

REFERENCES

[1] V. Potdar, A. Sharif and E. Chang, "Wireless Sensor Networks: A Survey," Advanced Information Networking and Applications Workshops, Bradford, 26-29 May 2009, pp. 636-641.

[2] Ding, W. and Marchionini, G. 1997. A Study on Video Browsing Strategies. Technical Report. University of Maryland at College Park.

[3] J. Chen and C. Li, "Coverage Optimization Based on Improved NSGA-II in Wireless Sensor Network," IEEE International Conference on Integration Technology (ICIT), Shenzhen, 20-24 March 2007, pp. 614-618.

[4] X. Wang, S. Wang and D. W. Bi, "Dynamic Sensor Nodes Selection Strategy for Wireless Sensor Networks," 7th International Symposium on Communications and Information Technologies (ISCIT), Sydney, 16-19 October 2007, pp. 1137-1142.

[5] J. Weck,"Layout Optimization for a Wireless Sensor Network Using a Multi Objective Genetic Algorithm," IEEE 59th Vehicular Technology Conference, Milan, Vol.5, 2004, pp.2466-2470.

[6] L.C. Wei, C. W. Kang and J. H. Chen, "A Force-Driven Evolutionary Approach for Multi-Objective 3D Differentiated Sensor Network Deployment," IEEE 6th International Conference on Mobile Ad-hoc and Sensor Systems (MASS), Macau, 12-15 October 2009, pp. 983-988

[7] S. Dadwal, T.S. Panag, "Sensor Deployment Strategies in WSNs," 3rd National Conference on Recent Advances in Electronics and Communication Technologies, Ludhiana, 21-22 March 2013, pp 412-419.

[8] A. Howard, M. Mataric and G. Sukhatme, "Mobile Sensor Network Deployment Using Potential Fields: A Distributed, Scalable Solution to the Area Coverage Problem," The 6th International Symposium on Distributed Autonomous Robotics System, Fukuoka, 25-27 June 2002, pp. 299-308.

[9] Yi. Zou and K. Chakrabarty, "Sensor deployment and Target Localization Based on Virtual Forces," IEEE Societies Twenty Second Annual Joint Conference of the IEEE Computer and Communications (INFOCOM), San Fransisco, 1-3 April 2003, pp. 1293-1303

[10] M. R. Pac, A. M. Erkmen and I. Erkmen, "Scalable Self- Deployment of Mobile Sensor Networks: A Fluid Dynamics Approach," Proceedings of IEEE International Conference on Intelligent Robots and Systems (RSJ), Beijing, 9-15 October 2006, pp. 1446-1451.

[11] R.-S. Chang and S.-H. Wang, "Self-Deployment by Density Control in Sensor Networks," IEEE Transactions on Vehicular Technology, Vol. 57, No. 3, 2008, pp. 1745-755.

[12] G. Wang, G. H. Cao, T. F. Porta and W. S. Zhang, "Sensor Relocation in Mobile Sensor Networks," IEEE Societies 24th Annual Joint Conference of the IEEE Computer and Communications, Miami, 2005, pp. 2302-2312.

[13] M. Cardei, J. Wu, M. M. Lu and M. O. Pervaiz, "Maximum Network Lifetime in Wireless Sensor Networks with Adjustable Sensing Ranges, " IEEE International Conference on Wireless and Mobile Computing, Net- working and Communications, Montreal, 22-24 August 2005, pp. 438-445

[14] Y. Qu, S.V. Georgakopoulos, "An Average Distance Based Self-Relocation and Self-Healing Algorithm for Mobile Sensor Networks," Journal of Wireless Sensor Network, Vol. 4 No.11, pp. 257-263,2012.

[15] G. Wang, G. H. Cao and T. F. Porta, "Movement-Assisted Sensor Deployment," IEEE Transactions on Mobile Computing, Vol. 5, No. 6, 2006, pp. 640-652.

[16] W. H. Sheng, G. Tewolde and S. Ci, "Micro Mobile Ro- bots in Active Sensor Networks: Closing the Loop," Proceedings of IEEE International Conference on Intelligent Robots and Systems (RSJ), Beijing, 9-15 October 2006, pp. 1440-1445.
[17] Y. G. Mei, Y.-H. Lu, Y. C. Hu and C. S. G. Lee, "Energy Efficient Motion Planning for Mobile Robots," Proceedings of IEEE International Conference on Robotics and Automation (ICRA), New Orleans, 25 April-1 May 2004, pp. 4344-4349.
[18] M. K. Stojcev, M. R. Kosanovic and L. R. Golubovic, "Power Management and Energy Harvesting Techniques for Wireless Sensor Nodes," 9th International Conference on Telecommunication in Modern Satellite, Cable, and Broadcasting Services, 7-9 October 2009, pp. 65-72.
[19] S. Singh, E. Singla, "Optimal design of robotic arms in constrained environment," TEQIP-II sponsored National conference on 'Latest Developments in Materials, Manufacturing and Quality Control' MMQC, Bathinda , 18-19 March 2015.

AUTHORS

Shivani Dadwal had completed her B.Tech degree from RIEIT, Railmajra. Now she is pursuing her M.Tech from BBSEBC, Fatehgarh Sahib.

INTENTIONAL BLANK

EFFECT OF RIGIDITY ON TRILATERATION TECHNIQUE FOR LOCALIZATION IN WIRELESS SENSOR NETWORKS

Dr. Saroja Kanchi

Department of Computer Science,
1700 University Blvd,
Kettering University,
Flint MI 48504
skanchi@kettering.edu

ABSTRACT

The localization of wireless sensor networks is an important problem where the location of wireless sensors is determined using the distance between sensors. Trilateration is a geometric technique used to find location of points in 2D using distances. Using geometry, one can find the location of a point uniquely in 2D given its distance to three other points in 2D. The problem of finding the trilateration order of vertices even if the network of sensors is a uniquely localizabe is NP-Complete. The 2D localization problem is closely related to the problem of graph rigidity. A graph can be uniquely realized in 2D if and only if the underlying network graph is globally rigid. Therefore by examining the structure of the underlying graph for rigidity and localization guided by rigidity is another technique used in localization.

We study the performance of trilateration which is based on geometry and local information to see if it is effected by graph rigidity which is a global property. In particular, we compare the performance of the trilateration on connected non-rigid networks and connected rigid networks. We focus on sparse networks graphs of lower radius.

1. INTRODUCTION

The recent advancements in wireless communication and sensing technology have resulted in wide deployment of sensors in applications like environmental monitoring, search and rescue, military surveillance, and intelligent transportation, etc [1, 2, 3]. In these types of applications, the knowledge of the location of each sensor is important. Due to constraints of these application, however it is often difficult to preset the locations of sensors before they are deployed. Therefore, the capability of obtaining the positions of sensors after the deployment is fundamental to the success of the mission of sensor networks. Most of the node localization algorithms are based on range measurements, through either time of arrival (TOA) [4], time difference of arrival (TDOA) [5], or received signal strength (RSS) [6, 7]. The problem of *localization* is to derive the geolocation of a node given a set of known locations and range measurements to these locations. Given the available range measures, if there is only one position for the nodes in the network,

David C. Wyld et al. (Eds) : ACITY, DPPR, VLSI, WiMNET, AIAA, CNDC - 2015
pp. 205–212, 2015. © CS & IT-CSCP 2015 DOI : 10.5121/csit.2015.51318

then the network is localizable. Similarly, if a node has only one position that satisfies all the range measures relevant to it, it is localizable.

Yang [8] presented conditions for node localizability using Trilateration technique. Yang et al [9] presented a distributed algorithm for localization which uses an extended trilateration. Wheel graphs are used as basic localizable subgraph and localization then extends to adjacent wheels. Li [10] provided a path for the mobile beacon based on depth-first search and used a variation of trilateration in the DREAMS technique. The geometric technique of trilateration is attractive due to the fact that the algorithms can be implemented as distributed algorithms.

The problem of network localization is closely related to the graph rigidity. A network defined by a set of nodes and a set of known distances between the nodes can be localized only if the graph derived from the network is uniquely realizable. It has been shown [11, 12, 13] that for a graph to be uniquely realizable, it must be redundantly rigid and tri-connected. Jacobs [11] proposed a centralized polynomial algorithm to check the rigidity of a graph through pebble games. trilateration alone. However, these are centralized algorithms. The pebble game algorithm described in Section 2 is based on depth first search and therefore is not distributed. However, distributed algorithms are use- ful considering the nature in which the sensor networks are deployed in practical setting. Very often there will not be a central node that all nodes in the network can communicate with. Moreover in cases where nodes may move and find new distances to their neighbors, these new distances need not be updated to central location if a distributed algorithm is used for rigidity finding and hence localization. However rigidity is a global property of a graph and it is quite a challenge to check the rigidity property without a huge message complexity.

In this paper we study if Trilateration performs better on rigid graphs versus random graph. Note that if a graph is 6-connected it is globally rigid and therefore localizable. We deal with graphs of low radius and low connectivity to examine if the Trilateration performs better in graphs that are weakly rigid.

2. GRAPH RIGIDITY AND ITS RELATION TO WSN LOCALIZATION

In this section, we are going to introduce the theory in network localizability and rigidity. A detailed description can be found in [11, 12, 13]. Let a framework $p(G)$ be a graph G along with a mapping $p : V \rightarrow R^2$ which assigns each vertex to a point in the plane. A *finite flexing* of a framework $p(G)$ is a family of realizations of G, parameterized by t so that the location, r_i , of each vertex i, is a differentiable function of t and $| r_i (t) - r_i (t)|^2$ is constant for every $(i, j) \in E$. Thinking of t as time, and differentiating the edge length constraints, we have

$$(u_i - u_i)(r_i - r_j) = 0 \text{ for every (i, j)} \in \text{E} \quad (1)$$

An assignment of velocities that satisfies Eq. 1 for a particular framework is an infinitesimal motion of that framework. Every framework has three trivial infinitesimal motion: two translations, and a rotation. If a framework has a nontrivial infinitesimal motion it is infinitesimally flexible. Otherwise it is infinitesimally rigid. Checking for whether a particular framework is rigid or not, can be determined from the property of the graph.

Let $G = \{V,E\}$ denote a network of vertices $V = \{1, 2, ..., n\}$ and for any edge $(i, j) \in E$, the distance between V_i and V_j is precisely known. The network localization problem is to determine the unique position of each node in the network given the positions of available beacons and the distance between each pair $(i, j) \in E$. If under the given constraints, there is only one position for each node, then the network is localizable. The network localization problem is closely related to the Euclidean graph realization problem, in which coordinates are assigned to vertices of a weighted graph such that the distance between coordinates assigned to nodes joined by an edge is equal to the weight of the edge.

For a two dimensional graph with n vertices, the positions of its vertices have $2n$ degrees of freedom, of which three are the rigid body motions. Therefore graph is rigid if there are $2n - 3$ constraints. If each edge adds an independent constraint, then $2n-3$ edges should be required to eliminate all non-rigid motions of the graph. Clearly, if any induced subgraph with n vertices has more than $2n-3$ edges then these edges cannot be independent, which leads to the following Laman theorem [14]:

Theorem 1 *The edges of a graph* G = {V,E} *are independent in two dimensions if and only if no subgraph* G′ = {V ′,E′} *has more than* 2n′ − 3 *edges, where* n′ *is the number of nodes in* G′.

Corollary 1 *A graph with* 2n − 3 *edges is generically rigid in two dimensions if and only if no subgraph* G′ *has more than* 2n′ − 3 *edges.*

Laman's theorem characterizes generic rigidity. However, a direct implementation of it leads to a poor exponential algorithm. An efficient approach to check for rigidity is proposed in [11] based on a pebble game. Jacob *et. al* proposed Jacob's approach uses the following formulation of Laman algorithm:

Theorem 2 *[11] For a graph* G = {V,E} *having* m *edges and* n *vertices, the following are equivalent.*

- *The edges of* G *are independent in two dimensions.*
- *For each edge* (a, b) *in* G*, the graph formed by adding three additional edges identical to* (a, b) *has no induced subgraph* G′ *in which* m′ > 2n′ .

The basic idea behind Jacob's algorithm is to grow a maximal set *S* of independent edges one at a time. Initially, *S* is empty. Let's denote these basis edges by *E*. A new edge is added to *S* if it is discovered to be independent of the edges existing in *S*.

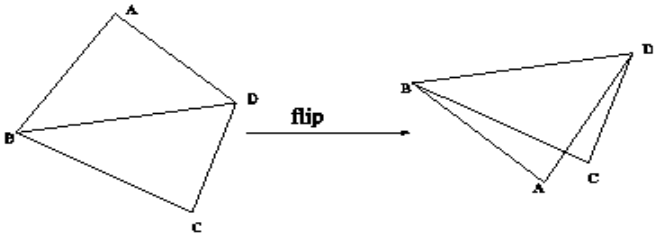


Figure 1: A generically rigid graph subject to flapping transformation. The two realizations are not continuous in two dimension space in that the second one is obtained by a flapping of the first one.

To check whether an edge e is independent of edges in S, each vertex is assigned two pebbles initially and a temporary set S' is created. S' contains all the edges in S plus four copies of e. The pebbles can only travel via the edges in S'. If all edges in S' can be covered by the pebbles, then we know that e is independent of all edges in S and e is added into S. This process is repeated until no more edges can be added into S. Then S is a maximal set of independent edges. If S contains $2n - 3$ edges, then the graph is generically rigid.

Having $2n - 3$ independent edges ensures the generic rigidity of a graph. However, it does not guarantee the unique realization of the network. A discontinuous change to the positions of nodes may lead to another realization which satisfies all the constraints of the network, as shown in Figure 1. The following theorem states the condition for a network to be uniquely realizable.

Theorem 3 *[13] A graph G with $n \geq 4$ vertices is uniquely realizable in two dimensions if and only if it is redundantly rigid and tri-connected.*

Redundant rigidity means after removing any single edge, the remaining graph is still generically rigid. A tri-connected graph is a connected graph such that deleting any two vertices (and incident edges) results in a graph that is still connected. When a network satisfies the condition in Theorem 3 can be uniquely localized given at least three nonlinear beacons in a two dimensional space.

3. TRILATERATION TECHNIQUE FOR RIGID GRAPHS

As we have discussed above, a network has to be globally rigid to be localizable. In this section, we are going to discuss in detail the proposed approach. To get a measure of localizability of the network we use a modified trilateration, a well-known technique in localizability. The technique of trilateration is based on the fact that in 2D, the unique location of the node can be determined given the distances to three other nodes whose locations are already known.

In iterative trilateration, we perform trilateration with starting vertices from different geographical parts of the graph, and choose the trilateration graph that has maximum number of localizable nodes.

Algorithm 1

1. Generate a graph that has flip rigidity
2. Repeat the following steps up to K times
3. Choose three starting vertices as anchors
4. Use trilateration to annex other vertices that connected to Anchor vertices
5. Mark the newly annexed vertices as anchors
6. Repeat annexation of vertices until no more vertices can be annexed
7. Count the number of annexed nodes.
8. Go back to step 2 with three other starting vertices.

4. RESULTS OF SIMULATION

We demonstrate that for sparse graphs, there is almost no difference in the performance of trilateration whether the graph is rigid or not. We implement a simulation on Matlab of multiple instances of graphs with 200 nodes over a ground of 100 by 100 with various radii. Figures 2 and 3 show random and rigid network of 200 nodes respectively.

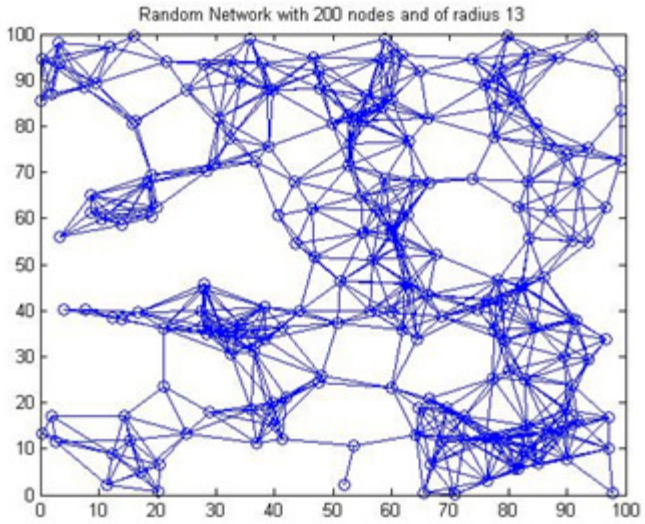


Figure 2: A random network of 200 nodes.

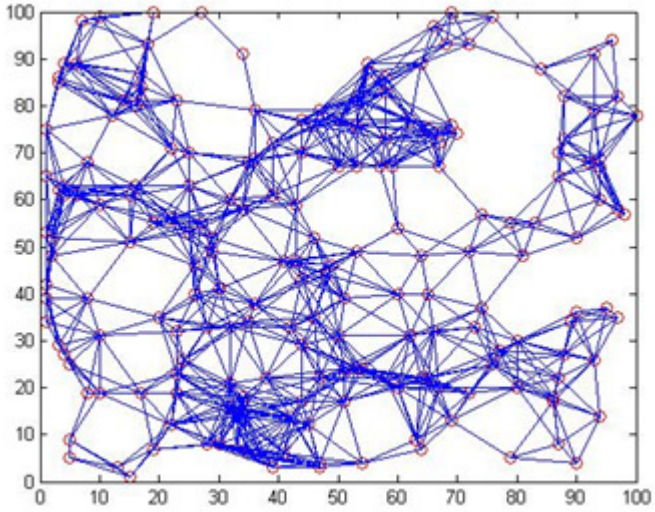


Figure 3: A rigid network of 200 nodes

The simulation is performed for multiple instances of the same radius for both random network and rigid network. We use Pebble game algorithm to check if a graph is rigid. Figure 4 and Figure 5 demonstrate the number of nodes localized for random and rigid networks with iterative trilateration using 100 instances of graphs for the averaging. Figures 6 and 7 show the number of edges used in localization of the corresponding network.

5. CONCLUSION

It can be seen that even though rigid networks in general are more localizable, for sparse graphs, only generic rigidity is a possibility and generic rigidity can help only marginally with localization. This is because rigidity is a global property of a graph but trilateration is a local property of a graph. Unless the global property can be translated into a collection of local properties, trilateration is not significantly helped by localization.

This leads us believe that perhaps two different approaches to localization should be considered. For sparse graphs, a localization that is based on finding localizable subgraphs and for dense graphs global rigidity can be used.

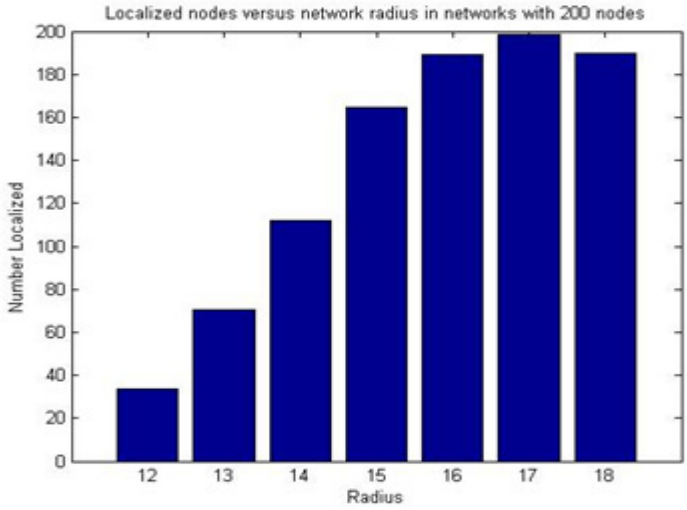


Figure 4: Number of nodes localized in a rigid network of 200 nodes.

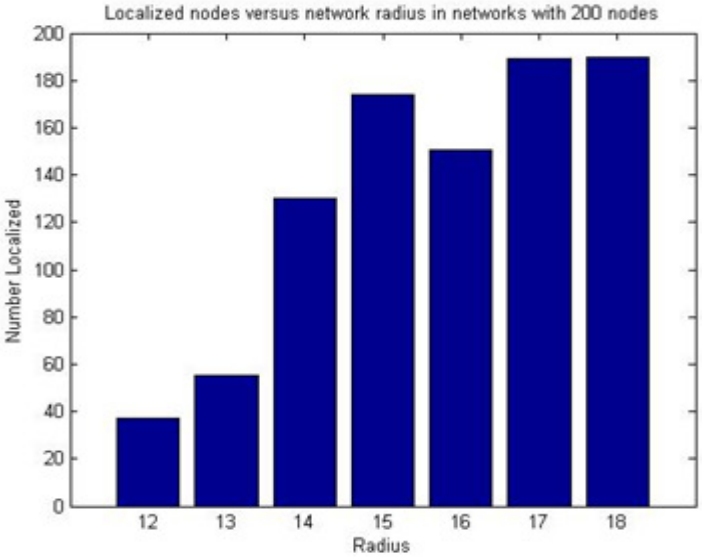


Figure 5: Number of nodes localized in a random network of 200 nodes.

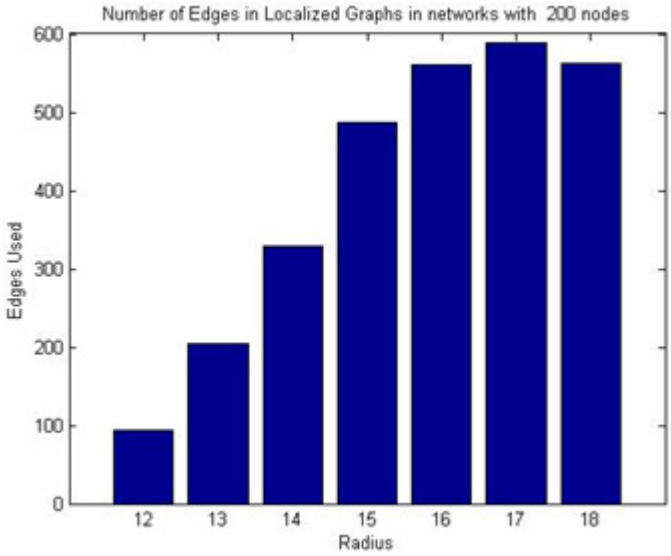


Figure 6: Number of edges used in localization of a rigid network of 200 nodes.

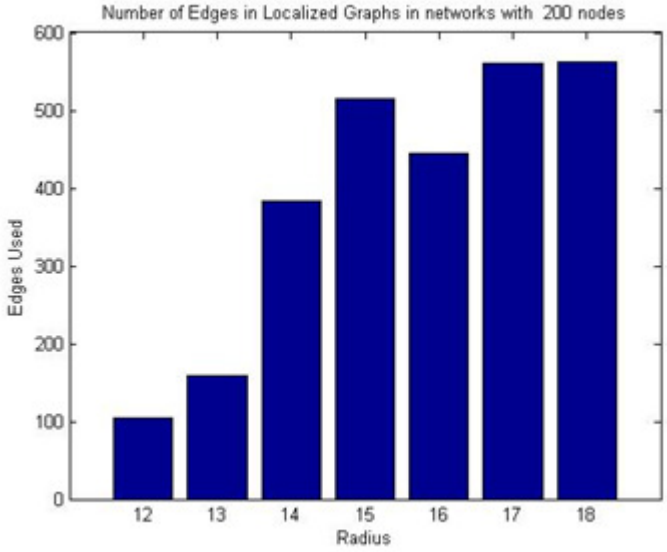


Figure 7: Number of edges used in localization of a rigid network of 200 nodes

REFERENCES

[1] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. A survey on sensor networks. IEEE Commun. Mag., 40:102–114, 2002.

[2] A. Mainwaring, J. Polastre, R. Szewczyh, D. Culler, and J. Anderson. Wireless sensor networks for habitat monitoring. In Proceedings of Wireless Sensor Network and Applications, 2002.

[3] S. N. Simic and S. Sastry. Distributed environmental monitoring using random sensor networks. In Proceedings of the 2nd International Workshop on Information Processing in Sensor Networks, pages 582–592, 2003.

[4] F. Zhao and L. Guibas. Wireless Sensor Networks: An Information Processing Approach. Elsevier and Morgan Kaufmann Publishers, 2004.

[5] C. Savarese, J.M. Rabaey, and J. Beutel, Locationing distributed ad hoc wireless sensor networks. In Proc. 2001 Int'l Conf. Acoustics, Speech, and Signal Processing (ICASSP 2001), volume 4, pages 2037–2040, May 2001.

[6] N. Bulusu, J. Heidemann, and D. Estrin. GPS-less low cost outdoor localization for very small devices. Technical Report 00-729, Computer science department ,University of Southern California, Los Angles, CA, 2000.

[7] X. Nguyen, M.I. Jordan, and B. Sinopli. A kernel-based learning approach to ad hoc sensor network localization. ACM Transactions on Sensor Networks, 1(1):134–152, 2005.

[8] Zheng Yang and Yunhao Liu. Understanding node localizability of wire- less ad hoc and sensor networks. IEEE TRANSACTIONS ON MOBILE COMPUTING, pages 1249–1260, 8 2012.

[9] Yunhao Liu Zheng Yang and Xiang-Yang Li. Beyond trilateration: On the localizability of wireless ad-hoc networks. In INFOCOMM, 2009.

[10] Isabelle Simplot-Ryl Xu Li, Nathalie Mitton and David Simplot-Ryl. Mobile-beacon assisted sensor localization with dynamic beacon mobility scheduling. 8th IEEE International Conference on Mobile Ad-hoc and Sensor Systems, 2011.

[11] Donald J. Jacobs and Bruce Hendrickson. An algorithm for two-dimensional rigidity percolation: the pebble game. Journal of Computation Physics, 137:346–365, 1997.

[12] T. Eren, O.K. Goldenberg, W. Whiteley, A.S. Yang, Y.R.; Morse, B.D.O. Anderson, and P.N Belhumeur. Rigidity, computation, and randomizationin network localization. In Proceedings of IEEE INFOCOM 2004, Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies, volume 4, pages 2673 – 2684, 2004.

[13] D. Goldenberg, A. Krishnamurthy, W. Maness, R. Yang, A. Young, and A. Savvides. Network localization in partially localizable networks. In Proceedings of INFOCOM 2005, 2005.

[14] G. Laman. On graphs and rigidity of plane skeletal structures. Journal of Engineering Mathematics, pages 331–340, 4 1970.

AUTHOR INDEX