

EASTERN

COMPUTER SCIENCE

Few Newer Perspectives

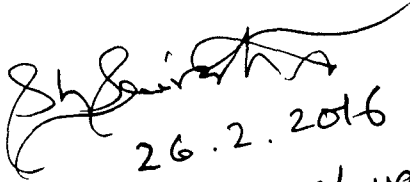


26310

- M. N. Bhattacharjee
- Smt. A. M. Mitri
- Banteilang Mukhim

Released on 26.02.2016.

With best wishes,


26.2.2016
Prof. S.K. Srivastava
V. C. NEHU, Shillong

Computer Science Few Newer Perspectives

Editors

Smt. A. M. Mitri
Shri Banteilang Mukhim
Dr. M. N. Bhattacharjee



EBH Publishers (India)
Guwahati-1

Smt. A. M. Mitri, Shri Banteilang Mukhim, Dr. M. N. Bhattacharjee
Computer Science – Few Newer Perspectives

SHILLONG COLLEGE	
LIBRARY	
Acc. No.	26310
Call No.	004.03 / MIT.
Price:	750/-
Date	5/3/16

All rights reserved. No part of this work may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior written permission of the copyright owner and the publisher.

The views expressed in this book are those of the Author, and not necessarily that of the publisher. The publisher is not responsible for the views of the Author and authenticity of the data, in any way whatsoever.

ISBN : 978-93-83252-80-0

© Author, 2015

First Published in 2015 by
EBH Publishers (India)
an imprint of Eastern Book House
136, M.L. Nehru Road, Panbazar
Guwahati-781 001, Assam (India)

Phone : +91 361 2513876, 2519231, 92070 45352

Fax : +91 361 2519231

Email : easternbookhouse@gmail.com.

www.easternbookhouse.in

Digitally Printed at : Replika Press Pvt. Ltd.

Printed in India



Boyce Road, Shillong – 793 003

www.shillongcollege.ac.in
Phone: 0364-2224903

shillcoll@yahoo.co.in
Fax: 0364-2502143

A NAAC Re-accredited Institution with CGPA 2.92 in 2010

Dr. K. D. Ramsiej
Principal,
Shillong College, Shillong

Foreword

It is a great pleasure for me to be associated with this important ISBN publication of the College, entitled '**Computer Science – Few Newer Perspectives**', which is the result of tiring efforts of the faculty members of the Computer Science and Applications Department of the College.

This book is mainly the outcome of the proceedings of a National Seminar on "Emerging Trends in Networking and Cloud Computing", held in the College in 2014. The various articles contained are the contributions of the participants who are academicians, scientists, eminent personalities etc. from across the country. These articles throw light on various issues with regard to the status of recent advances and current approaches in computer technology, internet facilities and related areas.

The initiatives taken by the teachers of the Computer Science and Applications Department of the College, and the IQAC of Shillong College to organise the National Seminar and then compiling the papers to publish the book deserve appreciation as it will certainly contribute to aware and arouse more interest in the field by sharing the knowledge and information and thus contributing more towards the currently most developing aspects of computer science and its applications.

I congratulate the faculty of Computer Science and Application Department of the College for their untiring efforts at bringing out this publication. I also thank the Coordinator of IQAC of the College for his keen interest and contribution for the success of the Seminar and bringing out this volume. I am sure that the contents of the book will serve the purpose, and the efforts of teachers will bear fruit.

Dated, Shillong
10th August, 2015

K. D. Ramsiej

Preface

With the dramatic advances of information and communication technologies, networking services are expected to integrate heterogeneous wire-line/wireless access technologies and the Internet backbone for providing various types of applications to both mobile and stationary users. The area generates interests in wireless sensor networks, vehicular networks, mobile communication networks, and the applications of networking technologies to cloud computing, including "Internet of Things" and "Cyber Physical System".

Networking and cloud computing are rapidly growing fields where innovations are happening very frequently. Hence there is an urge to understand the recent changes in networking and cloud computing. Recently these two topics make a wonderful rendezvous and produce motivational insights to the future research and industries.

Businesses are relying on cloud computing services for long term benefits including long term reduction of software and computer costs, improved data security (secure off site backup and storage) and increased functionality and customer service. As cloud computing continues to evolve and offer additional products and services, many businesses that are currently on the fence will make the decision to venture to "the cloud".

Recently, wireless sensor networks are used in a wide range of applications, including battlefield surveillance, health care, environment monitoring, building and bridge monitoring, and so on. The deployment, re-deployment, network self-cavity and abrupt diagnosis, self-reconfiguration, data collection with charge balance, query, data integration, and other popular topics with numerous changes of parameters, such as mobility, dynamic adjustment of sensing range and communication range, target coverage, barrier coverage, area coverage, wireless transmission mode (on land light wave or submarine acoustic wave), the arrangement of sleep/wake-up mechanism, the extensions of sensing ability like the visual camera sensor networks, the direction sensing capabilities, and so on, lead the research topics on the wireless sensor network to become rich and diverse in the future. The future research topics include: Information brokerage systems, Effective data aggregation, Underwater wireless sensor networks etc. With the extensive deployment of 3G networks and the standardization of 4G technologies (including WiMAX and LTE), the demand of real-time multimedia multicast applications are increased day by day. How to support such highly bandwidth-demanding applications in an efficient way has become a very challenging problem.

Whereas there are number of information technology based seminars/workshops being organized in different parts of our nation and the region, a seminar on the proposed topic is few and far between. It is an important emerging area which is generating sufficient interest because of its huge potential. A seminar or forum would provide platform for further knowledge as well as help entrepreneurship development.

The Seminar on 'Emerging Trends on Advanced Networking and Cloud Computing' was intended to apprise all stakeholders about the utility of cloud computing and advanced networking. It also provided the participants with the specialized skills required to understand existing network technology and how this is extended to cloud services. Participants shared their experiences on understanding of cloud computing fundamentals and applications including the risks, benefits and challenges of moving to the cloud. It was a learning experience in which the participants explored and evaluated the role of cloud resources in increasing competitive advantage. It gave a platform for enhancing and exchanging knowledge of experts from academia, R&D and industry and helps in providing in depth knowledge in the emerging development of advanced networking and cloud computing. It is an important emerging area and hence is expected to generate interest enhancing scope for further knowledge as well as help entrepreneurship development.

It is important that the knowledge resources be shared with all for the benefit of mankind which will encourage others for further development in the area. Developments in computer technology would be never ending, and consequently newer techniques and features would continue to emerge. The contents of this book would surely have some contribution in this regard.

Smt. A. M. Mitri

Shri Banteilang Mukhim

Dr. M. N. Bhattacharjee



Contents

<i>Foreword</i>	–	iii
<i>Preface</i>	–	v–vi
<i>List of Figures</i>	–	ix–xi
<i>List of Tables</i>	–	xiii
<i>List of Contributors</i>	–	xv–xvi
1. A Survey of Cloud Computing Strategies – <i>Arindam Roy, Sunita Sarkar, Bipul S Purkayastha</i>	–	1–14
2. Security Issues in Cloud Computing – <i>Surmila Thokchom</i>	–	15–24
3. Network Security: Challenges and Future Prospects – <i>Dr. Prodipto Das</i>	–	25–33
4. Roaming Agreements Simplified: A Supplementary Benefit of Enhancing Subscribers' Anonymity in Mobile Networks – <i>Hiten Choudhury</i>	–	34–44
5. Wireless Sensor Network Performance Analysis through Ad-Hoc Routing Protocols using Qualnet – <i>Hrituparna Paul, Dr. Prodipto Das</i>	–	45–54
6. Neural Network Based Modified AODV Routing Protocol in VANET – <i>Soumen Saha, Soumen Roy, Dr. Utpal Roy and Dr. Devadutta Sinha</i>	–	55–80
7. Recent Development and Challenges In Mobile Ad Hoc Network – <i>Indrani Das, Bipul Shyam Purkayastha and Sanjoy Das</i>	–	81–84
8. Cyber Space And Its Theories – <i>Arjun Chetry and Moutan Sarkar</i>	–	85–90
9. Cloud Computing Scenario in Indian SMEs: An Empirical Study – <i>Sanjeev Kumar Singh and Bipul Syam Purkayastha</i>	–	91–98
10. Proposing a Solution to Dynamic Mac Address Change of a Virtual Machine – <i>Partha Pratim Das, Rini Elis Tirkey, Jyotismita Talukdar and Ankit Singh</i>	–	99–109

11. Categorization of Traffic Hijacking Techniques in Cloud Ecosystem – <i>Narayan Tripathi, Pallavi Yadav and Prateek Chaturvedi</i>	– 110–118
12. Networking Aspects in Cloud Computing – <i>Bharathi Paleti, Vishesh Shukla and Uttkarsh Upmanyu</i>	– 119–126
13. Collaborative Recommender Systems via User-Item Subgroups – <i>Sharmistha Kakati and Sarat Kr. Chettri</i>	– 127–137
14. Application of First Order Fuzzy Time Series In Enrollment Forecasting Technique – <i>Jutang P. Swer and B. Borah</i>	– 138–145
15. Analysis of Distributed Data with Preserved Privacy – <i>Sarat Kumar Chettri</i>	– 146–157
16. Sentiment Analysis of Twitter Data Base on Hashtag – <i>Stevenson Mawa and N Donald Jefferson Thabah</i>	– 158–164
Index	– 165–167

List of Figures

<i>F. No.</i>	<i>Title of the Figure</i>	<i>Pages</i>
1.1 :	The Grid	– 3
1.2 :	The Cloud	– 3
1.3 :	Components of Cloud Computing Architecture	– 4
1.4 :	Cloud Infrastructure Components	– 5
1.5 :	Issues Handled by Cloud Infrastructure	– 6
1.6 :	Cloud Computing stack showing three distinct categories within Cloud Computing	– 7
2.1 :	Architecture of Cloud Computing	– 16
4.1 :	Authentication and key agreement procedure	– 37
4.2 :	AV generation	– 38
4.3 :	Current trust model	– 40
4.4 :	Simplified trust model	– 41
5.1 :	Random Waypoint Model	– 47
5.2 :	D view of scenario simulation in Qualnet	– 51
5.3 :	Throughput	– 52
5.4 :	Average End to End Delay(s)	– 52
5.5 :	AverageJitter(s)	– 53
6.1 :	Block diagram of the fuzzy neural network for Modified AODV (M-AODV) Routing Protocol	– 63
6.2 :	Testing Scenario	– 67
6.3 :	Number of Packet in collision in AODV vs aiAODV	– 69
6.4 :	Number of Packet drop in AODV vs aiMAODV	– 69
6.5 :	In throughput of AODV vs aiMAODV	– 70
6.6 :	Out throughput of AODV vs aiMAODV	– 70
6.7 :	Number of Packet in collision in AODV vs aiAODV	– 71
6.8 :	Number of Packet drop in AODV vs aiMAODV	– 71

<i>F No.</i>	<i>Title of the Figure</i>	<i>Pages</i>
6.9 :	In throughput of AODV vs aiMAODV	– 72
6.10 :	Out throughput of AODV vs aiMAODV	– 72
6.11 :	Number of Packet in collision in AODV vs aiAODV	– 73
6.12 :	Number of Packet drop in AODV vs aiMAODV	– 73
6.13 :	In throughput of AODV vs aiMAODV	– 74
6.14 :	Out throughput of AODV vs aiMAODV	– 74
6.15 :	Number of Packet in collision in AODV vs aiAODV	– 75
6.16 :	Number of Packet drop in AODV vs aiMAODV	– 75
6.17 :	In throughput of AODV vs aiMAODV	– 76
6.18 :	Out throughput of AODV vs aiMAODV	– 76
6.19 :	Number of Packet in collision in AODV vs aiAODV	– 77
6.20 :	Number of Packet drop in AODV vs aiMAODV	– 77
6.21 :	In throughput of AODV vs aiMAODV	– 78
6.22 :	Out throughput of AODV vs aiMAO	– 78
10.1 :	Types of Hypervisors	– 100
10.2 :	VMware Workstation	– 101
10.3 :	SUSE Linux Enterprise Desktop	– 102
10.4 :	DHCP	– 103
10.4 :	Sample of a dhcpd.conf File	–105–106
10.5 :	Reverse ARP (RARP)	– 108
10.6 :	Separate IP MAC address duplicity of any two virtual machines	– 109
11.1 :	Illustring different Traffics	– 111
11.2 :	CAGR Survey Report on Internet Traffic	– 112
11.3 :	Growth of Traffic	– 113
11.4 :	Distribution of Attack Techniques in Cloud Computing	– 115
11.5 :	BENEFITS of Cloud Computing	– 116
12.1 :	Network Architecture of Public Cloud	– 121
12.2 :	Network Architecture of Private Cloud	– 122
12.3 :	Network Architecture of Hybrid Cloud	– 123
13.1 :	Block diagram of the proposed system based on user	– 130

<i>F. No.</i>	<i>Title of the Figure</i>	<i>Pages</i>
15.1 :	Stepwise graphical representation of MDAV2k method where dots represents record, arrow represent distance measurement and triangle represents centroid ($k = 3$)	– 152
15.2 :	Information Loss and Disclosure Risk Comparison for Different k Values for Tarragona and Census Datasets	– 155
15.3 :	Information Loss and Disclosure Risk Comparison for Different k Values for EIA Dataset	– 155
16.1 :	Percentage of tweets	– 162
16.2 :	Total numbers of tweets which fall on different categories	– 162

List of Tables

<i>T. No.</i>	<i>Title of the Table</i>	<i>Page</i>
1.1 :	Estimated costs of infrastructure for the application servers, two database Servers and a load balancer across internal, managed and Cloud deployment models. Source O'Reilly Media: George Reese	– 12
5.1 :	Scenario Description	– 51
6.1 :	Data set generation	– 67
6.2 :	AODV testing parameters	– 68
6.3 :	aiAODV testing parameters	– 68
14.1 :	Actual Enrollment and Forecasted Enrollment according to the proposed method	–143–144
14.2 :	A comparison of the Enrollment forecasting results of different forecasting methods using First Order	– 144
15.1 :	Original Dataset D	– 149
15.2 :	Vertical Partitioning of Dataset D between Two Different Sites: (a) Site S1; (b) Site S2	– 149
16.1 :	Tweets description	– 160
16.2 :	Confusion Matrix	– 161
16.3 :	Total number of tweets	– 162
16.3 :	Most occurrence item.	– 163

List of Contributors

1. **Ankit Singh**, School of Technology, University of Technology and Management, Shillong, India.
2. **Arindam Roy**, Department of Computer Science, Assam University, Silchar, Assam-788011, India.
3. **Arjun Chetry**, Computer Programmer, M-tech 2nd year North Eastern Police Academy, Umsaw, Ri-Bhoi, Meghalaya Dept of CSE, affiliated to JNTU, Hyderabad.
4. **B. Borah**, Department of Computer Science & Engineering, Tezpur University, Tezpur - 784028, Assam.
5. **Bharathi Paleti**, School of Technology, University of Technology and Management, Shillong, Meghalaya-793003.
6. **Bipul Shyam Purkayastha**, Department of Computer Science, Assam University, Silchar, Assam-788011, India.
7. **Devadutta Sinha**, Department of CSE, University of Calcutta, Kolkata, WB, India.
8. **Hiten Choudhury**, Dept. of Computer Science, St. Edmund's College, Shillong, Meghalaya.
9. **Hrituparna Paul**, Assistant Professor, CSE, Department. NIT, Agartala, Tripura.
10. **Indrani Das**, Research Scholar, Jawaharlal Nehru University, New Delhi.
11. **Jutang P. Swer**, Department of Computer Science & Engineering, Tezpur University, Tezpur - 784028, Assam (Presently Department of Computer Science and Applications, Shillong College, Shillong).
12. **Jyotisma Talukdar**, School of Technology, University of Technology and Management, Shillong, India.
13. **Moutan Sarkar**, Computer Programmer, M-tech 2nd year North Eastern Police Academy, Umsaw, Ri-Bhoi, Meghalaya Dept of CSE, affiliated to JNTU, Hyderabad.
14. **N Donald Jefferson Thabab**, Department of Computer Science, Shillong College, Shillong, India
15. **Narayan Tripathi**, School of Technology, University of Technology and Management, Shillong, India.
16. **Pallavi Yadav**, School of Technology, University of Technology and Management, Shillong, India.

17. **Partha Pratim Das**, School of Technology, University of Technology and Management, Shillong, India.
18. **Prateek Chaturvedi**, School of Technology, University of Technology and Management, Shillong, India.
19. **Prodipto Das**, Assistant Professor, Department of Computer Science, Assam University, Silchar, Assam.
20. **Rini Elis Tirkey**, School of Technology, University of Technology and Management, Shillong, India.
21. **Utpal Roy**, Department of Computer & System Sciences, Siksha-Bhavana, Visva-Bharati, WB, India.
22. **Uttkarsh Upmanyu**, School of Technology, University of Technology and Management, Shillong, Meghalaya-793003.
23. **Vishesh Shukla**, School of Technology, University of Technology and Management, Shillong, Meghalaya-793003.
24. **Sanjoy Das**, Galgotias University, Greater Nodia, U.P.
25. **Sanjeev Kumar Singh**, Department of Mathematics, Union Christian College, Shillong, Meghalaya-793122, India
26. **Sarat Kr. Chettri**, Department of Computer Science & Engineering and Information Technology, School of Technology, Assam Don Bosco University, Guwahati, Assam.
27. **Sharmistha Kakati**, Department of Computer Science & Engineering and Information Technology, School of Technology, Assam Don Bosco University, Guwahati, Assam.
28. **Soumen Saha**, Department of CSE, University of Calcutta, Kolkata, WB, India.
29. **Soumen Roy**, Department of CSE, University of Calcutta, Kolkata, WB, India.
30. **Sunita Sarkar**, Department of Computer Science, Assam University, Silchar, Assam-788011, India.
31. **Surmila Thokchom**, Department of CSE, NIT Meghalaya.
32. **Stevenson Mawa**, Department of Computer Science, Shillong College, Shillong, India.

A Survey of Cloud Computing Strategies

Arindam Roy

Sunita Sarkar

Bipul S Purkayastha

Abstract

Cloud Computing is typically defined as a type of computing that relies on sharing computing resources rather than having local servers or personal devices to handle applications. A number of big Multinational Companies are into the business of Cloud Computing these days-Google Cloud, Amazon Cloud and Microsoft Azure - to name a few. The Economics of Cloud Computing has been evolving over the years and this is bound to attract industry-large, medium or small- into embracing Cloud Computing in the coming years.

Keywords: *Cloud Computing, Google Cloud, Microsoft Azure*

Introduction

Cloud computing has become the new buzz word driven largely by marketing and service offerings from big corporate players like Google, IBM and Amazon. Cloud computing is the next stage in evolution of the Internet. Cloud computing provides the means through which everything - from computing power to computing infrastructure, applications, business processes etc can be delivered to a client as a service wherever and whenever it is needed. The generally accepted definition of Cloud Computing comes from the National Institute of Standards and Technology (NIST) [1]. The NIST definition runs to several hundred words [2] but essentially says that: Cloud computing is a model for

enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. What this means in plain terms is the ability for end users to utilize parts of bulk resources and also these resources can be acquired quickly and easily.

NIST also cites several characteristics that it sees as essential for a service to be considered "Cloud". These characteristics include;

- On-demand self-service. The ability for an end user to sign up and receive services without the long delays that have characterized traditional IT
- Broad network access. Ability to access the service via standard platforms (desktop, laptop, mobile etc)
- Resource pooling. Resources are pooled across multiple customers [3]
- Rapid elasticity. Capability can scale to cope with demand peaks [4]
- Measured Service. Billing is metered and delivered as a utility service [5]

The rest of the Paper is organized as follows:- Section 2 describes in brief the evolution and history of cloud computing , Section 3 describes the architecture of cloud computing ,Section 4 illustrates the economics of cloud computing and Section 5 concludes the Paper.

Evolution and History of Cloud Computing

The mainframe and terminal application was the first type of client-server application that was very popular. At that time, storage and CPU were very expensive and the mainframe pooled both types of resources and served them to thin-client terminals. With the advent of the PC revolution, which brought huge storage and cheap CPUs to the average desktop, the file server gained in popularity as a means to enable document sharing and archiving. The file server served up storage resources to clients, while the CPU cycles needed to do productive work with those resources which were all produced and consumed within the confines of the PC client.

In the early 1990s, the budding Internet finally had enough computers attached to it that academics and researchers began thinking seriously about how to connect those machines together to create massive, shared pools of storage and compute power that would be much larger than what any one institution could afford to build. This is when the idea of "**the grid**" began to take shape.

The term "grid" is misinterpreted as synonyms for "cloud" as they both are made up of a large collection of inter connected computers. However they are two different things.

Grid computing requires the use of software that can divide a program as one large system image into several thousand computers. One problem with grid is that if one piece of the software on a node fails, other pieces of the software on other nodes may fail. This is alleviated if that component has a failover component on another node, but problems can still arise if components rely on other pieces of software to accomplish one or more grid computing tasks. Large system images and to operate their associated hardware and maintain them can contribute entail large capital and operating expenses.

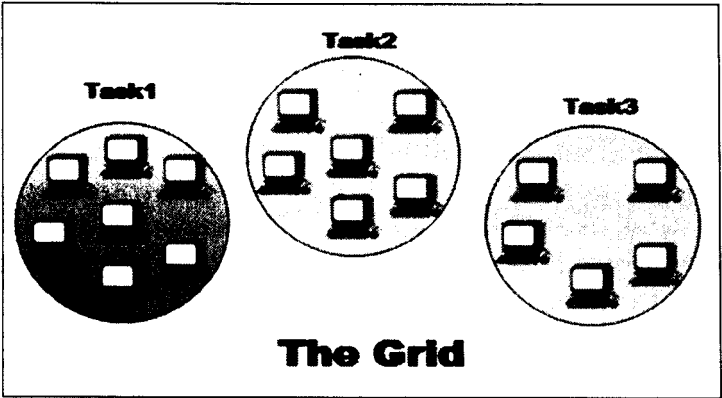


Fig.1.1: The Grid

Cloud computing evolves from grid computing and provides on-demand resource provisioning. Grid computing may or may not be in the cloud depending on what type of users are using it. If the users are systems administrators and integrators, they would be concerned how things are maintained in the cloud. They upgrade, install, and virtualize servers and applications. If the users are consumers, they do not care how things are run in the system.

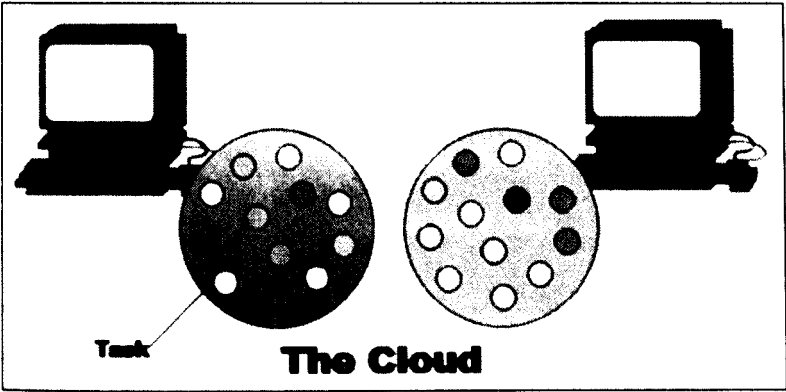


Fig. 1.2: The Cloud

The difference between grid computing and cloud computing is hard to grasp because they are not always mutually exclusive. In fact, they are both used to economize computing by maximising existing resources. However, the difference between the two lies in the way the tasks are computed in each respective environment. In a computational grid, one large job is divided into many small portions and executed on multiple machines. This characteristic is fundamental to a grid; not so in a cloud. The computing cloud is intended to allow the user to avail of various services without investing in the underlying architecture. Cloud services include the delivery of software, infrastructure, and storage over the Internet (either as separate components or as a complete platform) based on user demand.

Architecture of Cloud Computing

The Cloud Computing architecture comprises of many cloud components, each of them is loosely coupled. The cloud architecture can be broadly divided into two parts namely Front End and Back End. The Front End and Back End are connected by Internet.

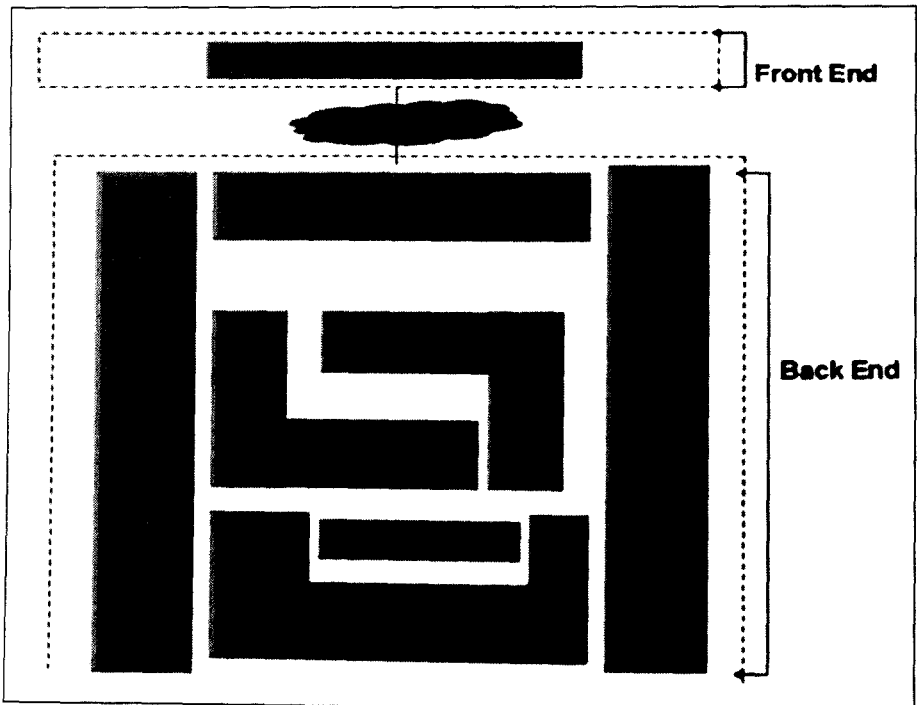


Fig. 1.3: Components of Cloud Computing Architecture

Front End refers to the client part of cloud computing system. It consists of interfaces and applications that are required to access the cloud computing platforms, e.g., Web Browser.

Back End refers to the cloud itself. It consists of all the resources required to provide cloud computing services. It comprises of huge data storage, virtual machines, security mechanism, services, deployment models, servers, etc.

Cloud Infrastructure Components

Cloud Infrastructure consists of servers, storage, network, management software, and deployment software and platform virtualization.

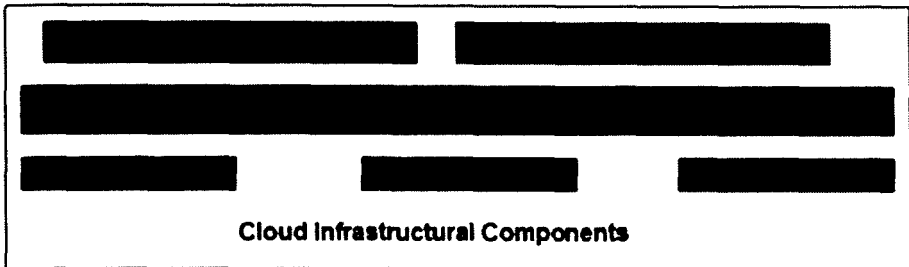


Fig. 1.4: Cloud Infrastructure Components

The Hypervisor is a firmware or low-level program that acts as a Virtual Machine Manager. It allows to share the single physical instance of cloud resources between several tenants. The Management Software helps to maintain and configure the infrastructure. The Deployment software helps to deploy and integrate the application on the cloud. Network is the key component of cloud infrastructure. It allows to connect cloud services over the Internet. It is also possible to deliver network as a utility over the Internet, i.e., the consumer can customize the network route and protocol.

Server helps to compute the resource sharing and offer other services such as resource allocation and de-allocation, monitoring resources, security, etc.

Cloud uses distributed file system for storage purpose. If one of the storage resources fails, then it can be extracted from another one which makes cloud computing more reliable.

Infrastructural Issues

There are some fundamental infrastructural issues which a cloud infrastructure should handle as shown in the following diagram:

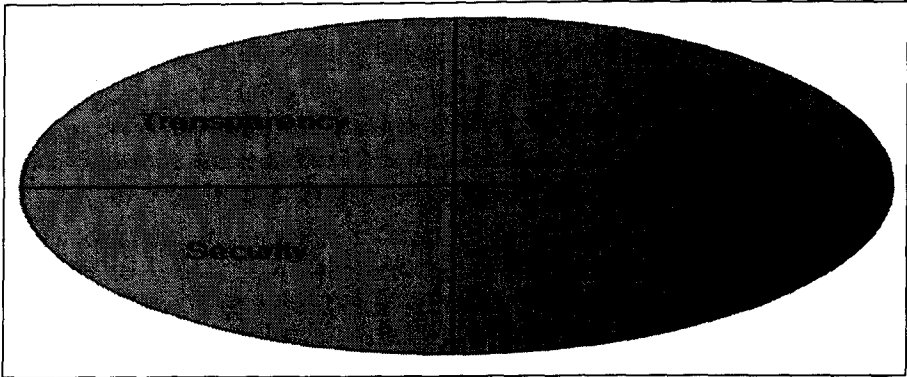


Fig. 1.5: Issues Handled by Cloud Infrastructure

- i. **Transparency:-** Virtualization is the key to share resources in cloud environment. But it is not possible to satisfy the demand with a single resource or server. Therefore, there must be transparency in resources, load balancing and application, so that we can scale them on demand.
- ii. **Scalability:-** Scaling up an application delivery solution is not that easy as scaling up an application because it involves configuration overhead or even re-architecting the network. So, application delivery solution is needed to be scalable which will require such virtual infrastructure where resource can be provisioned and de-provisioned easily.
- iii. **Intelligent Monitoring:-** To achieve transparency and scalability, application solution delivery will need to be capable of intelligent monitoring.
- iv. **Security:-** The mega data centre in the cloud should be securely architected. Also the control node, an entry point in mega data centre also needs to be secure.

Cloud Computing Stack

The diagram below depicts the Cloud Computing stack - it shows three distinct categories within Cloud Computing: Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS).

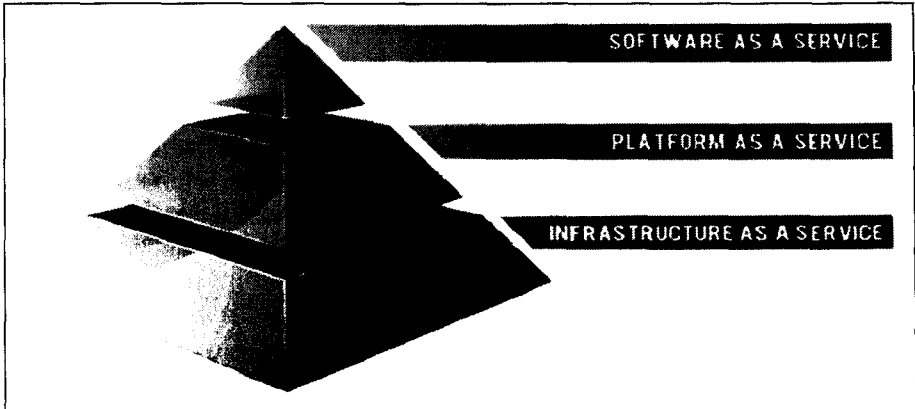


Fig. 1.6: Cloud Computing stack showing three distinct categories within Cloud Computing

- SaaS applications are designed for end-users, delivered over the web
- PaaS is the set of tools and services designed to make coding and deploying those applications quickly and efficiently.
- IaaS is the hardware and software that powers it all - servers, storage, networks, operating systems

Software as a Service

Software as a Service (SaaS) is defined as software that is deployed over the internet... [8] With SaaS, a provider licenses an application to customers either as a service on demand, through a subscription, in a "pay-as-you-go" model, or (increasingly) at no charge when there is opportunity to generate revenue from streams other than the user, such as from advertisement or user list sales. SaaS is a rapidly growing market as indicated in recent reports that predict ongoing double digit growth [9]. This rapid growth indicates that SaaS will soon become commonplace within every organization and hence it is important that buyers and users of technology understand what SaaS is and where it is not suitable.

- Characteristics of SaaS:** Like other forms of Cloud Computing, it is important to ensure that solutions sold as SaaS in fact comply with generally accepted definitions of Cloud Computing. Some defining characteristics of SaaS include:
 - Web access to commercial software
 - Software is managed from a central location
 - Software delivered in a "one to many" model

- Users not required to handle software upgrades and patches
 - Application Programming Interfaces (APIs) allow for integration between different pieces of software
- ii. **Where SaaS May Not be the Best Option:** While SaaS is a very valuable tool, there are certain situations where we believe it is not the best option for software delivery. Examples where SaaS may not be appropriate include;
- Applications where extremely fast processing of real time data is required
 - Applications where legislation or other regulation does not permit data being hosted externally
 - Applications where an existing on-premise solution fulfils all of the organization's needs

Software as a Service may be the best known aspect of Cloud Computing, but developers and organizations all around the world are leveraging Platform as a Service, which mixes the simplicity of SaaS with the power of IaaS, to great effect.

Platform as a Service

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

- i. **Characteristics of PaaS:** There are a number of different takes on what constitutes PaaS but some basic characteristics include [16];
- Services to develop, test, deploy, host and maintain applications in the same integrated development environment. All the varying services are needed to fulfill the application development process
 - Web based user interface creation tools help to create, modify, test and deploy different UI scenarios
 - Multi-tenant architecture where multiple concurrent users utilize the same development application
 - Built in scalability of deployed software including load balancing and failover
 - Integration with web services and databases via common standards

- Support for development team collaboration - some PaaS solutions include project planning and communication tools
 - Tools to handle billing and subscription management
- ii. **Where PaaS May Not be the Best Option:** In course of time PaaS will become the predominant approach towards software development. The ability to automate processes, use pre-defined components and building blocks and deploy automatically to production will provide sufficient value for it to be highly persuasive. That said, there are certain situations where PaaS may not be ideal, examples include;
- Where the application needs to be highly portable in terms of where it is hosted
 - Where proprietary languages or approaches would impact on the development process
 - Where a proprietary language would hinder later moves to another provider - concerns are raised about vendor lock-in [20]
 - Where application performance requires customization of the underlying hardware and software

Infrastructure as a Service

Infrastructure as a Service (IaaS) is a way of delivering Cloud Computing infrastructure - servers, storage, network and operating systems - as an on-demand service. Rather than purchasing servers, software, datacenter space or network equipment, clients instead buy those resources as a fully outsourced service on demand [23].

Generally IaaS can be obtained as public or private infrastructure or a combination of the two. "Public cloud" is considered as infrastructure that consists of shared resources, deployed on a self-service basis over the Internet. By contrast, "private cloud" is infrastructure that emulates some of Cloud Computing features, like virtualization, but does the same on a private network. Additionally, some hosting providers are beginning to offer a combination of traditional dedicated hosting alongside public and/ or private cloud networks. This combination approach is generally called "Hybrid Cloud".

- i. **Characteristics of IaaS:** As with the two previous sections, SaaS and PaaS. IaaS is a rapidly developing field. But there are some core characteristics which describe what IaaS is. IaaS is generally accepted to comply with the following:
- Resources are distributed as a service
 - Allows for dynamic scaling

- Has a variable cost, utility pricing model
- Generally includes multiple users on a single piece of hardware

There are a plethora of IaaS providers out there from the largest Cloud players like Amazon Web Services [22] and Rackspace [21].

As mentioned previously, the line between PaaS and IaaS is becoming more blurred as vendors introduce tools as part of IaaS that help in deployment including the ability to deploy multiple types of clouds [20].

ii. **Where IaaS May Not be the Best Option:** While IaaS provides massive advantages for situations where scalability and quick provisioning are beneficial, there are situations where its limitations may be problematic. Examples of situations where we would advise caution with regard to IaaS include;

- Where regulatory compliance makes the off shoring or outsourcing of data storage and processing difficult.
- Where the highest levels of performance are required, and on-premise or dedicated hosted infrastructure has the capacity to meet the organization's needs.

Economics of Cloud Computing

The 80-20 rule is often used within organizations to illustrate the large effects that small variables can have. It was first suggested by business management thinker Joseph Juran and originally called the Pareto principle¹ after Italian economist Vilfredo Pareto. Rather than being an absolute measure, it tends to be a generalization that is intended to make a point about distribution curves. The most well known use of the rule is the sales 80-20 rule which says that 80% of revenue for a business is derived from 20% of customers.

Gartner's findings show that Information Technology has its own series of 80-20 rules. IT maintenance accounts for around 80% of total IT expenditure. [3]

Cloud Computing is a force that helps flip this ratio and gives IT departments the ability to spend 80% of their time on core business processes, like business application design. It's for this reason, the ability to go from 20% of time and money dedicated to core business processes to 80%, that the economics of Cloud Computing is so compelling. Nowhere is the current model's inefficiency more evident than in the opportunity costs that organizations pay to manage their own computing needs.

Opportunity cost, a concept first developed by British philosopher John Stuart Mill, is a basic economic premise that is concerned with the costs related

to the choices NOT made by someone. Opportunity cost is: "the cost related to the next-best choice available to someone who has picked among several mutually exclusive choices. It is a key concept in economics. Opportunity costs are not restricted to monetary or financial costs: the real cost of output forgone, lost time, pleasure or any other benefit that provides utility should also be considered opportunity costs [4].

With this explanation of opportunity cost, we can now apply the concept to a decision to either retain on-premise IT or move to the Cloud. As we've already seen, roughly 80% of IT time and expenditure is wasted on processes that don't create any value for the organization (beyond maintaining the status quo). The opportunity cost of not choosing the Cloud is therefore the benefit that can accrue to the organization through optimal utilization of that 80%. To put it simplistically, a move to the Cloud can make the difference between an organization being 20% efficient, and one being 80% efficient.

While opportunity cost, and the value to be gained by reducing that cost, is a compelling benefit of moving to Cloud Computing, many critical readers will want to see more concrete examples of the economics at work. To this end it is important to understand the gains to be made from a move away from capital expenditure, and over to operating expenditure.

Traditional IT expenditure has been very capital intensive. Hardware had to be bought outright and software licenses were generally an expenditure that appeared on the balance sheet. For this reason the decision making process for technology spend became very drawn out.

One of the core tenets of Cloud Computing is that it is a recurring expenditure model and as we detailed in a previous report⁵ it is much like telephone or electricity expenditure which is accounted for as a standard operating expense.

There are several distinct reasons that OpEx, or operating expenditure, is preferred to CapEx, or capital expenditure.

OpEx is beneficial for the organization, as it gives it the flexibility to terminate costs at will. With a capital purchase, the server or software being acquired is fully committed to. Regardless of whether it is being utilized, the ongoing costs (by way of depreciation or financing costs) still need to be borne. Contrast this with OpEx where, in the event that the item is no longer required, payments can cease rapidly. It is for this reason that many companies prefer leased vehicles in place of purchasing them outright.

Table 1.1: Estimated costs of infrastructure for the application servers, two database. Servers and a load balancer across internal, managed and Cloud deployment models. Source O'Reilly Media: George Reese [7].

	Internal IT	Managed Services	Cloud
Capital Investments			
Set up costs	\$40,000	\$0	\$0
Monthly Services	\$1,000	\$5000	\$1,000
Monthly labour	\$0	\$4000	\$2,400
Cost over 3 years	\$3,200	\$0	\$1,000
Savings Gained	\$1,49,000	\$1,29,000	\$1,06,000
	0%	13%	29%

But big companies have embraced the cloud more slowly than its fans had expected. IDC, a research firm, estimates that businesses will spend \$100 billion on cloud computing over the coming years. But it is a fraction of the \$2 trillion or so that companies will spend on information technology (IT). Some are holding back because their IT teams claim they can run things more cheaply internally. Others are wary of entrusting sensitive data to another firm's servers.

Such corporate reluctance may soon be dissipated, as price cuts make cloud computing cheaper still. At the end of March 2014 Google slashed prices by between 30% and 85% on cloud services such as application processing and data storage. The move aimed at boosting Google's own cloud-computing business-sparked a swift response from AWS, which cut some prices by up to 65%. Microsoft, which is also determined to be big in the cloud, followed suit with cuts of its own.

Conclusion

Cloud Computing is a term that doesn't describe a single thing - rather it is a general term that spreads over a variety of services from Infrastructure as a Service at the base, through Platform as a Service as a development tool and through to Software as a Service replacing on-premise applications.

For organizations looking to move to Cloud Computing, it is important to understand the different aspects of Cloud Computing and to assess their own situation and decide which types of solutions are appropriate for their unique needs.

Cloud Computing is a rapidly accelerating revolution within IT and will become the default method of IT delivery moving into the future, not least, because of simply the economics of Cloud Computing.

References

1. <http://csrc.nist.gov/groups/SNS/cloud-computing/>
2. <http://csrc.nist.gov/groups/SNS/cloud-computing/cloud-def-v15.doc>
3. Virtualization - The ability to increase computing efficiency http://broadcast.rackspace.com/hosting_knowledge/whitepapers/Revolution_Not_Evolution-Whitepaper
4. Scalability and fast provisioning - for IT at web scale - http://broadcast.rackspace.com/hosting_knowledge/whitepapers/Revolution_Not_Evolution-Whitepaper.pdf
5. From Water-wheel to Utility Power - An analogy for the Cloud - http://broadcast.rackspace.com/hosting_knowledge/whitepapers/Revolution_Not_Evolution-Whitepaper.pdf
6. <http://www.qrimp.com/blog/blog.The-Difference-between-IaaS-and-PaaS.html>
7. <http://m.zdnet.com/blog/forrester/is-the-iaaspaas-line-beginning-to-blur/583>
8. http://en.wikipedia.org/wiki/Software_as_a_service
9. <http://www.readwriteweb.com/cloud/2010/07/sass-providers-challenge-the-k.php>, <http://www.networkworld.com/news/2010/101810-saas-on-a-tear-says.html>
10. <http://www.salesforce.com>
11. <http://www.networkworld.com/news/2010/050610-gartner-saas-adoption-on-the.html>
12. <http://www.zendesk.com/blog/groupon-defenders-of-the-customer-experience>
13. <http://www.groupon.com>
14. <http://www.zendesk.com>
15. <http://www.zendesk.com/blog/hey-groupon-thanks-a-million>
16. http://en.wikipedia.org/wiki/Platform_as_a_service and <http://java.dzone.com/articles/what-platformservice-paas>
17. <http://code.google.com/appengine/>
18. <http://www.microsoft.com/windowsazure/>
19. <http://www.salesforce.com/platform/>
20. <http://www.zdnet.com/blog/saas/cogheads-demise-highlights-paas-lock-out-risk/668>
21. <http://www.menumate.com/>
22. <http://trineo.co.nz/>
23. <http://diversity.net.nz/wp-content/uploads/2011/01/Moving-to-the-Clouds.pdf>
24. <http://diversity.net.nz/wp-content/uploads/2011/01/Moving-to-the-Clouds.pdf>
25. <http://aws.amazon.com/>
26. <http://www.rackspace.com/>
27. <http://m.zdnet.com/blog/forrester/is-the-iaaspaas-line-beginning-to-blur/583>
28. <http://www.beyonddiet.com/BD/Categories/Fat-Burning-Foods>

29. http://www.tutorialspoint.com/cloud_computing/cloud_computing_infrastructure.htm
30. <https://www.google.co.in/webhp?sourceid=chrome-instant&ion=1&espv=2&ie=UTF-8#q=iaas%20paas%20saas>
31. http://www.rackspace.com/knowledge_center/whitepaper/understanding-the-cloud-computing-stack-saas-paas-iaas

Security Issues in Cloud Computing

Surmila Thokchom

Abstract

Security of Cloud computing is a very important issue in the face of popularity it has gain globally. There are number of economic benefits of Cloud Computing but at the same time it requires defence mechanism. Cloud computing becomes a successful and popular business model due to its charming features. The people whose concern is the cloud security continue to hesitate to transfer their business to cloud. The paper takes note of some of the security measures.

Introduction

Cloud computing is a technology that is gaining in popularity every day. It represents a new business model and computing paradigm, which enables on-demand provisioning of computational and storage resources. Economic benefits consist of the main drive for cloud computing due to the fact that cloud computing offers an effective way to reduce capital expenditure and operational expenditure. Cloud computing can be defined as : "A large-scale distributed computing paradigm that is driven by economies of scale, in which a pool of abstracted, virtualized, dynamically-scalable, managed computing power, storage, platforms, and services are delivered on demand to external customers over the Internet.

Cloud Architecture

Fig. 2.1 depicts the general architecture of a cloud platform, which is also called cloud stack. Building upon hardware facilities (usually supported by modern data centers), cloud services may be offered in various forms from the bottom

layer to top layer. In the cloud stack, each layer represents one service model. Infrastructure-as-a-Service (IaaS) is offered in the bottom layer, where resources are aggregated and managed physically (e.g., Emulab) or virtually (e.g., Amazon EC2), and services are delivered in forms of storage (e.g., GoogleFS), network (e.g., Openflow), or computational capability (e.g., HadoopMapReduce). The middle layer delivers Platform-as-a-Service (PaaS), in which services are provided as an environment for programming (e.g., Django) or software execution (e.g., Google App Engine). Software as a Service (SaaS) locates in the top layer, in which a cloud provider further confines client flexibility by merely offering software applications as a service.

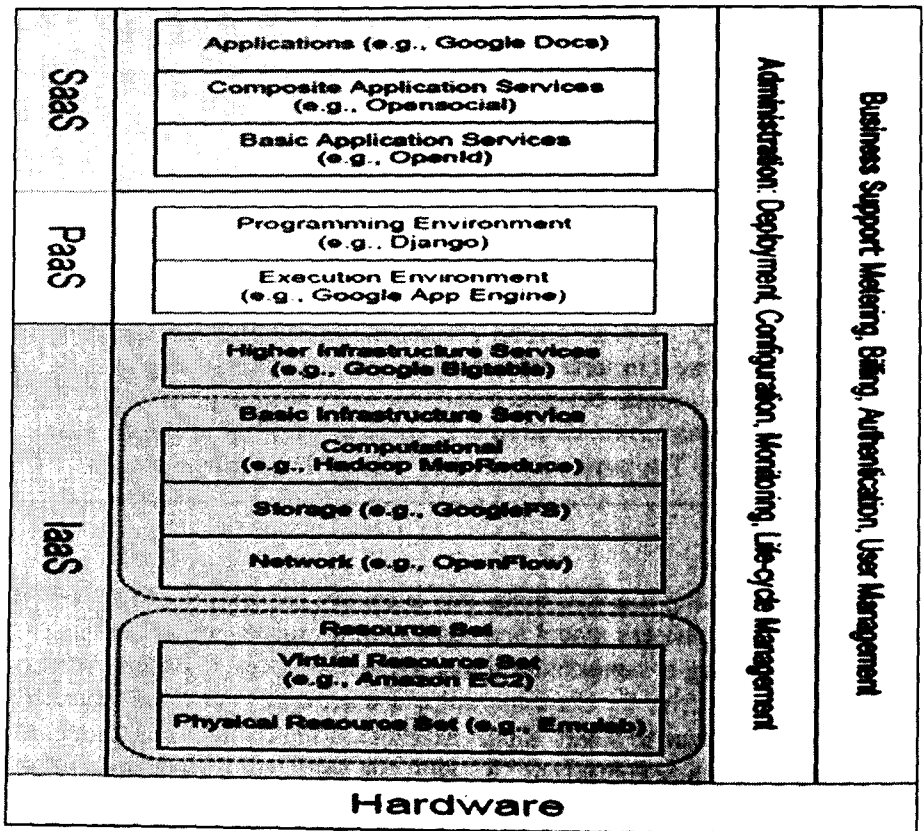


Fig. 2.1: Architecture of Cloud Computing

Cloud Characteristics

The Cloud Security Alliance has summarized five essential characteristics that illustrate the relation to, and differences from, traditional computing paradigm.

- Where regulatory compliance makes the off shoring or outsourcing of data stor
- On-demand self-service- A cloud customer may unilaterally obtain computing capabilities, like the usage of various servers and network storage, as on demand, without interacting with the cloud provider.
- Broad network access- Services are delivered across the Internet via a standard mechanism that allows customers to access the services through heterogeneous thin or thick client tools (e.g., PCs, mobile phones, and PDAs).
- Resource pooling- The cloud provider employs a multi-tenant model to serve multiple customers by pooling computing resources, which are different physical and virtual resources dynamically assigned or reassigned according to customer demand. Examples of resources include storage, processing, memory, network bandwidth, and virtual machines.
- Rapid elasticity- Capabilities may be rapidly and elastically provisioned in order to quickly scale out or rapidly released to quickly scale in. From customers' point of view, the available capabilities should appear to be unlimited and have the ability to be purchased in any quantity at any time.
- Measured service -The service purchased by customers can be quantified and measured. For both the provider and customers, resource usage will be monitored, controlled, metered, and reported.

Security Challenges

Cloud computing becomes a successful and popular business model due to its charming features. The people whose concern is the cloud security continue to hesitate to transfer their business to cloud. Security issues have been the dominate barrier of the development and widespread use of cloud computing. There are three main challenges for building a secure and trustworthy cloud system:

- **Outsourcing-** Outsourcing brings down both capital expenditure and operational expenditure for cloud customers. However, outsourcing also means that customers physically lose control on their data and tasks. The loss of control problem has become one of the root causes of cloud insecurity. To address outsourcing security issues, first, the cloud provider shall be trustworthy by providing trust and secure computing and data storage; second, outsourced data and computation shall be verifiable to customers in terms of confidentiality, integrity, and other security services.
- **Multi-tenancy-** Multi-tenancy means that the cloud platform is shared

and utilized by multiple customers. Moreover, in a virtualized environment, data belonging to different customers may be placed on the same physical machine by certain resource allocation policy. Adversaries who may also be legitimate cloud customers may exploit the co-residence issue. A series of security issues such as data breach computation breach flooding, etc., are incurred. Although Multi-tenancy is a definite choice of cloud vendors due to its economic efficiency, it provides new vulnerabilities to the cloud platform. Without changing the multi-tenancy paradigm, it is imperative to design new security mechanisms to deal with the potential risks

- **Massive data and intense computation-** cloud computing is capable of handling mass data storage and intense computing tasks. Therefore, traditional security mechanisms may not suffice due to unbearable computation or communication overhead. For example, to verify the integrity of data that is remotely stored, it is impractical to hash the entire data set. To this end, new strategies and protocols are expected.

Cloud Confidentiality

When dealing with cloud environments, confidentiality implies that a customer's data and computation tasks are to be kept confidential from both the cloud provider and other customers. Confidentiality remains as one of the greatest concerns with regards to cloud computing. This is largely due to the fact that customers outsource their data and computation tasks on cloud servers, which are controlled and managed by potentially untrustworthy cloud providers.

Threats to Cloud Confidentiality

1. **Cross-VM attack:** A cross-VM attack via side channel [3] exploits the nature of multi-tenancy, which enables that VMs belonging to different customers may co-reside on the same physical machine.
2. **Malicious System Administrator:** Privilege system administrator of the cloud provider can perform attacks by accessing the memory of the customer's VMs. Xenaccess enables a sysadmin to directly access the VM memory at run time by running a user level process in Domain0.

Defence Strategies

Some of the approaches to address cross-VM attack are discuss below.

1. **Placement prevention [3]:** Cloud provider may obfuscate co-residence by having dom0 not respond in tracer route and/or by randomly assigning internal IP addresses to launch VMs. To reduce the success rate of placement, cloud providers might let the users decide where to put their VMs.

2. Co-residency Detection: To ensure physical isolation, client should be enable to verify its VMs exclusive use of a physical machine. HomeAlone [7] is a system that detects co-residency by employing a side-channel as a detection tool.
3. No Hype [5] [6]: Virtualization provided by the hypervisor, is the key underlying technology in the cloud infrastructure. But the use of virtualization is the source of a significant security concern. NoHype architecture proposes at removing the hypervisor while still retaining the key features of virtualization such as arbitrating access to CPU, memory and I/O devices, acting as a network device, managing the starting and stopping of guest virtual machine.

Approaches to Address Malicious System Administrator

1. Trusted cloud computing platform (TCCP) [4]

TCCP offers a closed box execution environment for IaaS services. It guarantees confidential execution of guest virtual machines. The design goals of TCCP are to confine the VM execution inside the secure perimeter and other is a system administrator with root privileges is unable to access the memory of a VM hosted in a physical node.

2. Self-service Cloud Computing [1]

Cloud infrastructure rely on virtual machine monitors (VMMs) to flexibly administer and execute client VMs. VMMs implement a trusted computing base (TCB) that virtualizes the CPU, memory and I/O devices and manages VMs. In VMMs like Xen and Hyper-V, the TCB has two parts; the hypervisor and an administrative domain. The hypervisor directly control the physical hardware and runs at the highest processor privilege level. The administrative domain called Dom0, is a privilege VM that has the privilege to start/stop client VMs, change client VM configuration, monitor their physical resources utilization and perform I/O for virtualized devices. Providing Dom0 with such privileges leads to two problems. One is security and privacy of client VMs and the other is inflexible control over client VMs.

Self Service Cloud (SSC) Computing model addresses the above two problems. SSC privilege model reduces the power of the administrative domain and gives clients more flexible control over their own VMs. SSC's privilege model splits the responsibilities traditionally entrusted with dom0 between a new system-wide administrative domain (Sdom0) and per-user administrative domains (Udom0s), service domain (SDs) and mutually trusted service domains (MTSDs).

Udom0 (User dom0) is a per-user administrative domain that can monitor and control the set of VMs of a particular client. When a client attempts to start a VM in SSC, it is assigned its own Udom0 domain. This domain creates the user

VMs that perform the actual work for the client and it can also delegate its privileges to service domain (SDs), which are special purpose user domains that can perform privileged system services on UdomUs. Clients can leverage SDs to implement services such as memory introspection to verify VM integrity, intrusion detection and storage encryption.

Sdom0 (System dom0) is the system wide administrative domain which has the privilege to start/stop Udom0 domain upon request by client and to run drivers for virtualized devices. MTSDs execute privileged services that check regulatory compliance in a manner that is mutually agreed upon between cloud provider and the client.

Threats to Cloud Integrity

1. Data loss or manipulation: The cloud servers which are giving storage as a service in cloud storage are distrusted in terms of security and reliability. In others words data stored in cloud storage may be lost or modified maliciously or accidentally.
2. Dishonest computation in remote servers: Data computations are not transparent to customer in the cloud. So it is difficult for the customer to check if the cloud servers behave unfaithfully and return incorrect result or if some problem arises and uses outdated, vulnerable code etc.

Defence Strategies

1. Provable Data possession (PDP): PDP [10] requires that the data is pre-processed in the setup phase in order to leave some meta-data on the client side for verification purposes subsequently, for the data to be sent to the cloud server. Once the client feels a necessity to check the integrity at the later time, he sends a challenge to the cloud server, which will respond with a message based on the data content. After combining the reply and the local meta-data, the client is able to prove whether the integrity of the data is violated. PDP can detect server misbehaviour with low computational and storage overhead. PDP is applicable to static files.
2. Proof of Retrievability (POR): POR [11] employs an auditing protocol when solving a similar problem to PDP. The problem is that each of the two enables clients to check the integrity of outsourced data without having to retrieve it. PoR is designed to minimize the storage in client and server side, the communication complexity of an audit, and the number of data-blocks accessed during an audit. The POR protocol is the user stores only a key, which is used to encode the file F in order to get the encrypted file F' . The task is that a set of sentinel

values are embedded into F' , and the server only stores F' without knowing where the sentinels may be since the sentinels are indistinguishable from regular data blocks. In the challenge and response protocol, the server is asked to return a certain subset of sentinels in F' . If the server has tampered with or deleted F' , there is high probability that certain sentinels are also corrupted or lost; this causes the server to be unable to generate a complete proof for the original file. Therefore, a client has evidence to prove that the server has corrupted the file. Due to the fact that the number of sentinels is limited, POR adopts error-correcting codes to recover the file with only a small fraction being corrupted. Similar to PDP, PoR can only be applied to static files. A subtle change to the file will ruin the protocol and completely confuse the clients.

3. **Scalable PDP:** Scalable PDP [12] is an improved version of the original PDP. The difference in the two are described as the following: 1) scalable PDP adopts symmetric key encryption instead of public-key to reduce computation overhead, but scalable PDP does not support public very verification due to symmetric encryption; 2) scalable PDP has added dynamic operations on remote data. One limitation of scalable PDP is that all challenges and answers are pre-computed, and the number of updates is limited and fixed as a priori.
4. **Dynamic PDP:** The goal of Dynamic PDP (DPDP)[13] is to support full dynamic operations. The purpose of dynamic operations is to enable authenticated insert and delete functions with rank-based authenticated directories that are built on a skip list. The experiment result shows that, although the support of dynamic updates costs certain computational complexity, DPDP is practically efficient. The DPDP protocol introduces three new operations, which are known as PrepareUpdate, PerformUpdate and VerifyUpdate. PrepareUpdate is run by a client in order to generate an update request that includes the updates to perform.
5. **HAIL:** A High-Availability and Integrity Layer (HAIL)[14] for cloud storage differs from the prior work with regards to the fact that it considers a distributed setting in which a client must spread a file across multiple servers with redundancy and only store a small constant state in local machine. The main threats that HAIL combats are mobile adversaries, which may possibly corrupt file F by undermining multiple servers.
6. **Improved DPDP:** Improved DPDP [19] divides file into blocks, generates a tag for each block, computes a hash value of for each tag, uses tags to ensure the integrity of the file blocks and uses hash values

to ensure integrity of tags. Compared with original DPDP, improved DPDP reduces the computational and communication complexity from $(\log n)$ to constant. However client needs to store some secret values which may create some additional storage expense which is 0.02% of the original file size.

7. Iris: Iris[15] is a system that supports dynamic POR. In Iris, tenants obtain strong assurance on data integrity, data freshness and data retrievability in case of accidental or adversarial cloud failures. The main goal of Iris is to support integrity protection of both file system and meta -data and continuous verification of full file system correctness and availability with minimum overhead. In Iris, heavy caching on the enterprise side is strictly necessary. The reason for this is, firstly if local caching is not performed, the cost of network transfer to and from the cloud will far outweigh any storage costs savings. Secondly without local caching individual operation latency will be prohibitive for the file system to be usable.
8. Robust DPDP [17]: Robustness is a necessary property for all Remote data checking schemes. Robustness is usually achieved by integrating forward error-correcting codes with remote data checking. Reed-Solomon codes provide efficient error correction capabilities in the static case but their linear nature imposes a high communication cost when even a small portion of the original file needs to be updated. R-DPDP provides two construction, first is ?R-D construction, which achieves robustness by extending techniques from the static to the dynamic setting. This scheme is efficient in encoding but requires a high communication cost for updates. Second construction is VLCG (Variable Length Constraint Group) overcome the drawback of the ?R-D by decoupling the encoding for robustness from the position of symbols in the file and instead relying on the value of symbols and reducing expensive insert/delete operations to append/modify operating when updating the Reed-Solomon- coded parity data, which ensures efficient updates even under an adversarial setting.
9. DR-DPDP [18]: This system is a distributed and replicated DPDP which is transparent from the client's view point. In others words the cloud service provider (CSP) can flexibly manage its resources, perform its choice of load balancing and replication schemes in the background, while still providing provable storage for the client. Better performance since the load is distributed on multiple servers. Availability and reliability are also improved via replication. This system uses persistent rank-based authenticated skip lists to create centralized



and distributed variants of a dynamic version control system with optimal complexity.

Conclusion

In this paper, security and privacy issues in cloud computing are discussed. Also it has discussed the various existing defense mechanism dealing in the threats in cloud confidentiality and cloud data integrity.

References

1. Butt, S., Lagar-Cavilla, H.A., Srivastava, A., and Ganapathy, V. "Self-service cloud computing". In Proceedings of the 2012 ACM Conference on Computer and Communications Security. ACM Press, New York, 2012, 253-264.
2. Colp, P., Nanavati, M., Zhu, J., Aiello, W., Coker, G., Deegan, T., Loscocco, P., and Warfield, A. Breaking up is hard to do: Security and functionality in a commodity hypervisor. In Proceedings of the 23rd ACM Symposium on Operating Systems Principles. ACM Press, New York, 189-202.
3. T. Ristenpart, E. Tromer, H. Shacham, and S. Savage, "Hey, you, get off of my cloud: exploring information leakage in third-party compute clouds," Proc. 16th ACM conference on Computer and communications security, 2009, pp. 199-212.
4. N. Santos, K.P. Gummadi, and R. Rodrigues, "Towards trusted cloud computing," Proc. 2009 conference on Hot topics in cloud computing, 2009.
5. E. Keller, J. Szefer, J. Rexford, and R. B. Lee, "NoHype: virtualized cloud infrastructure without the virtualization," in Proc. 37th annual international symposium on Computer architecture, New York, NY, USA, 2010, pp. 350-361.
6. J. Szefer, E. Keller, R. B. Lee, and J. Rexford, "Eliminating the hypervisor attack surface for a more secure cloud," in Proc. 18th ACM conference on Computer and communications security, New York, NY, USA, 2011, pp. 401-412.
7. Yinqian Zhang, A. Juels, A. Oprea, and M. K. Reiter, "HomeAlone: Co-residency Detection in the Cloud via Side-Channel Analysis," in 2011 IEEE Symposium on Security and Privacy (SP), 2011, pp. 313-328.
8. Z Xiao, Y. Xiao, "Security and Privacy in cloud computing", IEEE Communications Surveys & Tutorials, Vol 15, No. 2, second Quarter 2013
9. M. Nanabati, P. Colp, B. Aiello, A. Warfield, "Cloud Security: A Gathering Storm", ACM communication, Vol. 57, No. 5, page 70-79, may 2014
10. G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, and D. Song, "Provable data possession at untrusted stores," In ACM CCS, pages 598-609, 2007.
11. A. Juels and B. S. Kaliski, "PORs: Proofs of retrievability for largefiles," In ACM CCS, pages 584-597, 2007.
12. G. Ateniese, R. D. Pietro, L. V. Mancini, and G. Tsudik, "Scalable and efficient provable data possession," SecureComm, 2008.
13. C. Erway, A. K'upc, "u, C. Papamanthou, and R. Tamassia, "Dynamic provable

- data possession," Proc. 16th ACM conference on Computer and communications security, 2009, pp. 213-222.
14. K.D. Bowers, A. Juels, and A. Oprea, "HAIL: A high-availability and integrity layer for cloud storage," Proc. 16th ACM conference on Computer and communications security, 2009, pp. 187-198.
 15. E. Stefanov, M. van Dijk, A. Oprea, and A. Juels, "Iris: A scalable cloud file system with efficient integrity checks," IACR ePrint Cryptography Archive, Tech. Rep. 2011/585, 2011.
 16. Q. Zheng and S. Xu, "Fair and dynamic proofs of retrievability," in Proc. of the ACM CODASPY '11, 2011.
 17. B. Chen, R. Curtmola, "Robust Dynamic provable data possession" ,IEEE, 2012, 32nd International Conference on Distributed Computing Systems Workshops.
 18. M. Etemad and A. Kupcu, "Transparent, Distributed and Replicated Dynamic Provable Data Possession, ACNS'13 Proceedings of the 11th International Conference on Applied Cryptography and Network Security
 19. F. Liu, D. Gu and H. Lu, "An Improved Dynamic Provable Data Possession Model", Proceedings of CCIS 2011.

Network Security: Challenges and Future Prospects

Prodipto Das

Abstract

Network security has become very important to personal computer users, organizations, and the military. With the advent of the internet, security became a major concern and the study of history of security allows a better understanding of the emergence of security technology. The internet structure itself allowed for many security threats to occur. The architecture of the internet, when modified, can reduce the possible attacks that can be sent across the network. Knowledge of the attack methods allows for the appropriate security to emerge. Many businesses secure themselves from the internet by means of firewalls and encryption mechanisms. The businesses create an intranet to remain connected to the internet but also remain secured from possible threats. The entire field of network security is vast and is in an evolutionary stage. The range of study encompasses a brief history dating back to internet's beginnings and the current development in network security. In order to understand the research being performed today, background knowledge of the internet, its vulnerabilities, attack methods through the internet, and security technology is important and therefore they are reviewed.

Introduction

The world is becoming more interconnected with the advent of the Internet and new networking technology. There is a large amount of personal,

commercial, military, and government information on networking infrastructures worldwide. Network security is becoming of great importance because of intellectual property that can be easily acquired through the internet.

There are currently two fundamentally different networks, data networks (connected servers) and synchronous network comprised of switches (backbone networks). The internet is considered a data network. Since the current data network consists of computer-based routers, information can be obtained by special programs, such as Trojan horses, planted in the routers. The synchronous network that consists of switches does not buffer data and therefore is not threatened by attackers. That is why security is emphasized in data networks, such as the internet, and other networks that link to the internet.

Why Network Security Is Necessary

Today, the Internet is made up of tens of thousands of networks, interconnected without boundary. Network security is essential in this environment because any organizational network is accessible from any computer in the world and, therefore, is potentially vulnerable to threats from individuals who do not require physical access to it. In a recent survey conducted by the Computer Security Institute (CSI), 70 percent of the organizations polled stated that their network security defences had been breached and that 60 percent of the incidents came from within the organizations themselves. While it is difficult to measure how many companies have had Internet-related security problems and the financial losses due to those problems, it is clear that the problems do exist.

Security Principles

There are many branches of security. If we consider the field of security as a hierarchy, we have security at the root and many branches leading outwards from that. For example, information security, and economic security may be considered subsets of the entire discipline of security. The field of security is concerned with protecting general assets. Information security is concerned with protecting information and information resources, such as books, faxes, computer data, and voice communication. Network security is concerned with protecting data, hardware and software on a computer network. For many years, the focus of security was on prevention. Every security technique and technology fall into at least one of the three elements: prevention, detection, and response. Security equation can be represented as:

$$\text{Protection} = \text{Prevention} + (\text{Detection} + \text{Response})$$

Understanding Principles of Network Security

Cryptography

Cryptography is a method of storing and transmitting data in a particular

form so that only those, for whom it is intended, can read and process it. Cryptography is closely related to the disciplines of cryptology and cryptanalysis. Cryptography includes techniques such as microdots, merging words with images, and other ways to hide information in storage or transit. However, in today's computer-centric world, cryptography is most often associated with scrambling plaintext (ordinary text, sometimes referred to as cleartext) into ciphertext (a process called encryption), then back again (known as decryption). Individuals who practise in this field are known as cryptographers [1].

Modern cryptography concerns itself with the following four objectives:

- 1) **Confidentiality** (the information cannot be understood by anyone for whom it was unintended)
- 2) **Integrity** (the information cannot be altered in storage or in transit between sender and intended receiver without the alteration being detected)
- 3) **Non-repudiation** (the creator/sender of the information cannot deny at a later stage his or her intentions in the creation or transmission of the information)
- 4) **Authentication** (the sender and receiver can confirm each other's identity and the origin/destination of the information)

Procedures and protocols that meet some or all of the above criteria are known as cryptosystems. Cryptosystems are often thought to refer only to mathematical procedures and computer programs; however, they also include the regulation of human behaviour, such as choosing hard-to-guess passwords, logging off unused systems, and not discussing sensitive procedures with outsiders.

Wireless Network Security

[A] Protecting the Confidentiality of Wireless Transmissions

Two types of countermeasures exist for reducing the risk of eavesdropping on wireless transmissions. The first involves methods for making it more difficult to locate and intercept the wireless signals. The second involves the use of encryption to preserve confidentiality even if the wireless signal is intercepted [2].

[B] Signal-Hiding Techniques

In order to intercept wireless transmissions, attackers first need to identify and locate wireless networks. There are, however, a number of steps that organizations can take to make it more difficult to locate their wireless access points. The easiest and least costly include the following: Turning off the service set identifier (SSID) broadcasting by wireless access points, Assigning cryptic names to SSIDs, Reducing signal strength to the lowest level that still provides requisite coverage or Locating wireless access points in the interior of the building,

away from windows and exterior walls. More effective, but also more costly methods for reducing or hiding signals include: Using directional antennas to constrain signal emanations within desired areas of coverage or Using signal emanation-shielding techniques, sometimes referred to as TEMPEST, 1 to block emanation of wireless signals.

[C] Encryption

The best method for protecting the confidentiality of information transmitted over wireless networks is to encrypt all wireless traffic. This is especially important for organizations subjected to regulations.

Challenges of Network Security [3][4]

1. State-sponsored espionage: This challenge highlights the need to protect critical data from politically or financially motivated threats. Critical data include the information needed to run network attached infrastructure as well as the intellectual property used to manage business and drive innovative solutions.
2. DDoS attacks: Security professionals in the financial services industry are likely to agree to our second challenge: monster DDoS attacks. We can expect to see a higher risk of business impacting threats with the shift from computer-based attacks generating large number of lower bandwidth events, to virtual server or cloud-based attacks generating ultra-high bandwidth events. With these new attack vectors it becomes even more beneficial to identify and mitigate large DDoS events while traffic is in the network cloud.
3. Password management: Here our challenge is - putting in place and enforcing stronger user-controlled passwords that are less likely to be broken. This educational and administrative challenge requires creative solutions and enforced policies. Or, we can look at alternatives in traditional passwords, such as the use of Federated ID's.
4. Sabotage: Sabotage of computer networks can affect critical infrastructure and ultimately impact corporate and backbone networks. This challenge is so perverse potentially because it combines social engineering with software based tools to provide a complex multi-vectorized attack profile.
5. Botnets: Botnets are everywhere. The challenge is that many botnet owners design systems that are more adaptive and redundant than many corporate and government networks. Controlling this agile attack vector - before it is used as an Advanced Persistent Threat (APT) which then migrates into smart mobile devices - is crucial.

6. Insider threat: A dissatisfied employee base provides a vector for insider security events, while the inadvertent injection of malware through removable media or Web interconnections can make any employee the origination point for a network security violation.
7. Mobility: Management and security of mobile networks and smart mobile devices become even more challenging when employees want to use their own devices for business purposes. The Bring Your Own Device (BYOD) trend exasperates this challenge when we look at protecting the critical information needed to manage the organization and the network without sacrificing the privacy of the employee's personal information and activities.
8. Internet: One of the greatest challenges to security professionals is the perception that the Internet, a best effort network, is a secure critical infrastructure. The Internet is an open connection of diverse networks. The challenge is to start treating critical networks as if they are critical to our operations. We need to put into effect policies that distinguish platforms and security levels based on business criticality. Control networks need different security level than general business communications. This includes using network embedded security controls to help reduce risks and to simplify security infrastructure.
9. Privacy laws: This final challenge is currently being legislated worldwide. We need to balance privacy with the need to gather information that can help address security breaches or fraud, while complying with associated legislation.

Present Trends of Network Security [5]

I. DDoS Attacks Get Sneaky

DDoS attackers will go from simple volumetric attacks to ones which take advantage of a site's specific performance characteristics. That's the prediction of security researchers at Neohapsis, a security and risk management consulting company.

DDoS attacks that intelligently target bottlenecks in performance, such as pages with a high server load (like database writes) or specific network bottlenecks (like login and session management), can magnify the impact of attacks which are simply volume-based and request the homepage of a site. So it's likely that we will begin to see the spread of tools which profile specific targets.

II. Insider Threats Remain Major Security Problem

According to a Cyber Security Watch survey insiders were found to be the cause in 21 per cent of security breaches, and a further 21 per cent may have

been due to the actions of insiders. More than half of respondents to another recent survey said it's more difficult today to detect and prevent insider attacks than it was in 2011, and 53 per cent of the respondents was increasing their security budgets in response to insider threats.

III. Security Worries Drive Cloud Consolidation

Organizations will look to buy more solutions from a single vendor and demand greater integration between solutions to automate security, according to Eric Chiu, president of HyTrust, a cloud security company. The fact that securing cloud environments is very different from securing traditional physical environments will drive greater consolidation in the market, he says.

IV. Legacy Systems Cause More Security Headaches

The spate of IT failures in banks and other high profile companies highlights a simple fact: Many of them are running legacy systems which are so old and out of date that they are becoming almost impossible to maintain. That's because there are few people with the skills and expert knowledge that would be needed to run them securely - even if they were updated to eliminate known vulnerabilities, which they frequently are not.

They often aren't updated because no one knows what impact that would have. It's inevitable that we'll see hackers going after such systems, exploiting vulnerabilities that can't easily be fixed.

V. Encryption Will Be Revisited

In the wake of revelations about the NSA, many companies are realizing that encryption may be the only thing that is protecting their data, and it may not be as strong as they imagined. What's more, if hackers are led to believe there is a weakness in a particular system - either accidental or intentional - they will pound on it until they find it.

As a result, many companies will look to improve the way they use encryption. The experts at Neohapsis advise to look for particular attention to be paid to cryptographic block modes like CBC and OFB, and authenticated modes like EAX, CCM and GCM. In addition to the encryption methods themselves, they also advise to look for insights and innovations around key management and forward security.

Types of Encryption [6]

[A] Symmetric Encryption

Symmetric encryption's job is to take readable data ("plaintext" in crypto parlance), scramble it to make it unreadable (protecting it from prying eyes while it's being stored on a disk or transmitted over a network), then unscramble it again when it's needed. It's generally fast, and there are lots of good encryption

methods to choose from. The most important thing to remember about symmetric encryption is that both sides-the encrypter, and the decrypter-need access to the same key.

A key, for symmetric encryption purposes, is a string of data that is fed to the encrypter in order to scramble the data and make it encrypted. It is best if this key is completely random, but there are ways to derive keys from (hopefully really good) passwords as well. The tricky part about using symmetric encryption is how to store the key and make it available only to the software that needs it.

[B] Asymmetric Encryption

Asymmetric encryption also takes readable data, scrambles it, and unscrambles it again at the other end, but there's a twist: a different key is used for each end. Encrypters use a public key to scramble the data, and decrypters use the matching private (secret) key on the other end to unscramble it again. The public key is literally just that, public; it can and should be published. (This is why asymmetric encryption is also often referred to as public-key cryptography.) But the private key must be kept private, protected much like the key for symmetric encryption. The good news is that this is easier, since only one party ever needs access to it: the party that needs to decrypt the messages. Some (but not all!) asymmetric encryption systems have one additional important capability: the ability to cryptographically sign data. In this system, the private key is used to make the signature, and the public key is used to verify it. You can thus prove, if you have data with a signature and the matching public key, that it was signed with the private key.

[C] Hashing

Hashing is what is actually happening when you hear about passwords being "encrypted". Strictly speaking, hashing is not a form of encryption, though it does use cryptography. Hashing takes data and creates a hash out of it, a string of data with three important properties:

- The same data will always produce the same hash
- It is impossible to reverse it back to the original data
- Given knowledge of only the hash, it's infeasible to create another string of data that will create the same hash (called a "collision" in crypto parlance)

Future Prospects of Network Security

What is going to drive the Internet security is the set of applications more than anything else. The future possibly is that the security will be similar to an immune system. The immune system fights off attacks and builds itself to fight

tougher enemies. Similarly, the network security will be able to function as an immune system. The trend towards biometrics could have taken place a while ago, but it seems that it isn't being actively pursued. Many security developments that are taking place are within the same set of security technology that is being used today with some minor adjustments.

Conclusion

Network security is an important field that is increasingly gaining attention as the internet expands. The security threats and internet protocol were analyzed to determine the necessary security technology. The security technology is mostly software based, but many common hardware devices are used. The current development in network security is not very impressive. Originally it was assumed that with the importance of the network security field, new approaches to security, both hardware and software, would be actively researched. It was a surprise to see most of the development taking place in the same technologies that are currently used. The embedded security of the new internet protocol IPv6 may provide many benefits to internet users. Although some security issues were observed, the IPv6 internet protocol seems to evade many of the current popular attacks. Combined use of IPv6 and security tools such as firewalls, intrusion detection, and authentication mechanisms will prove effective in guarding intellectual property for the near future. The network security field may have to evolve more rapidly to deal with the threats further into the future.

Acknowledgement

The author would like to thank the Organizing Secretary of National Seminar on Emerging Trends in Advanced Networking and Cloud Computing, 2014, Shillong College for his kind invitation to the author as a Resource Person. The author would like to thank also the Principal, Administrative Officer and all other staff members of Shillong College for giving the honour. Last but not the least the author would like to thank Prof. Bipul Shyam Purkayastha of Assam University, Silchar, for offering such opportunity to deliver the invited lecture.

References

1. W. Stallings, *Wireless and Mobile Systems*, Pearson India, 2010.
2. P. Das, *Improving the Quality of Service of Mobile Communication Network in Uncertain Areas*, Assam University Silchar, 2011.
3. R. Das, B. S. Purkayastha, P. Das and K. Ullah, "Simulation Study of Flooding Attack in the Mobile Ad Hoc Networks", *AU Journal of Science & Technology*, AUJST, Vol. VIII, Number II 2011, pp. 48-52. ISSN: 0975-2773.
4. P. Das, Lingfen Sun and B. S. Purkayastha, "Propagation Study during Rain on DTH Service in North East India", *AU International Journal of Science & Technology*, India, Vol. 7, No. 2, pp. 122-126, 2011. ISSN: 0975-2773

5. <http://www.esecurityplanet.com/network-security/7-security-trends-to-expect-in-2014.html>
6. <http://spin.atomicobject.com/2014/11/20/encryption-symmetric-asymmetric-hashing/>

4

Roaming Agreements Simplified: A Supplementary Benefit of Enhancing Subscribers' Anonymity in Mobile Networks

Hiten Choudhury

Abstract

In mobile networks, identity presentation should precede authentication and all other forms of security. Thus, the subscriber has to present his/her identity through an insecure channel, before any service can be extended to him/her. Ensuring anonymity to the subscriber is therefore a challenging security issue in mobile networks. Numerous research works have been carried out to strengthen the subscriber's anonymity in mobile networks. Early researches in this area were focused on protecting the subscribers' anonymity from an eavesdropper over the radio link between the user equipment and the visited network. Recent research however has gone a step ahead, by making an attempt to enhance the subscribers' anonymity by protecting the subscribers' identity from even the visited network that may practically be forged or compromised. This, as perceived by the author, would also bring in a supplementary benefit of simplifying 'roaming agreements', which is a pact between mobile operators. This pact enables a roaming subscriber of one operator to continue having access to the mobile services while in the service area of another operator. In this paper, an effort is made to elaborate on this additional benefit.

Key words: *Mobile Network, Identity Privacy, Network Operator*

Introduction

With the increase in availability and popularity of a range of mobile services, mobile devices are becoming an integral part of an individual. However, threats like location tracking and comprehensive profiling, where data about movement, usage, etc., of a subscriber is collected and linked to his/her identity to explore various attacks, have also emerged. Thus, identity privacy in mobile systems has become an important security issue.

Early researches in this area were focused on protecting the subscribers' anonymity from an eavesdropper over the radio link between the user equipment and the visited network. However, recent research has gone a step ahead, by making an attempt to enhance the subscriber's anonymity by protecting the subscriber's identity from even the visited network that may practically be forged or compromised. This, as perceived by the author, would also bring in a supplementary benefit of simplifying 'roaming agreements', which is a pact between mobile operators. This pact enables a roaming subscriber of one operator to continue having access to the mobile services while in the service area of another operator. In this paper, an effort is made to elaborate on this additional benefit.

The rest of the paper is organized as follows. In Section II, we explain the concept of identity privacy in global mobility networks. In Section III, we give an overview of the current status of identity privacy in global mobility networks with reference to a prominent network called Universal Mobile Telecommunications Network (UMTS). In Section IV, we discuss the identity privacy vulnerabilities in global mobility networks. In Section V, we discuss related work that envisages enhanced identity privacy in global mobility networks. In Section VI, we elaborate on how enhancing identity privacy may contribute in simplifying roaming agreements between operators. Finally, in Section VII, we conclude the paper.

Identity Privacy

Privacy has been a concern for people since the ancient times. Exposure of many of the activities such as movement, access to resources, usage behavior, etc., of a person may lead to his/her risk of physical security as well as security of his/her resources. One's activities may be revealed if his identity is known to the adversaries [1]. Hence, the confidentiality of one's identity is of paramount importance.

According to a recent press release of The World Bank, around three-quarters of the World's inhabitants now have access to a mobile phone [2] and the number is increasing with every passing day. These days, a subscriber uses a mobile phone to access variety of services including voice, rich communication services, and value added services. These services are used for making important

communications, accessing valuable resources, and for carrying out financial transactions, because of which a mobile phone is becoming an important tool for an individual's existence. Therefore, the need to protect an individual's identity that is used in a mobile system is as important as the need to protect other important personal identities like social security number and bank account numbers.

In mobile networks, each subscriber is registered with a home network. During registration, the subscriber is assigned a Subscriber Identity Module (SIM) that contains a unique and a permanent identity called the International Mobile Subscriber Identity (IMSI) that identifies the subscriber. The IMSI is a number (Eq. 1) that constitutes of a maximum of 15 decimal digits [3]. The rest 3 digits are the Mobile Country Code (MCC), which is followed by the Mobile Network Code (MNC) (either 2 digits, in case of European standard or 3 digits, in case of North American standard). The length of the MNC depends on the value of the MCC. The remaining digits are the Mobile Subscription Identification Number (MSIN) [4]. Thus, $IMSI = MCC||MNC||MSIN$ (1) Where, '||' denotes concatenation. The IMSI is used by the home network to uniquely identify each and every subscriber for authentication, authorization and billing purposes. The MCC identify the country of domicile of the mobile subscriber, whereas the MNC identify the home network of the mobile subscriber. The MSIN is used to uniquely identify a subscriber within the subscriber's home network.

Identity Privacy is considered a standard security requirement in any mobile telecommunication system [5] [6] [7]. The identity privacy of a subscriber is compromised if his/her permanent identity (i.e., the IMSI) is exposed to an adversary. Knowledge of the IMSI may allow an adversary to track and amass comprehensive profiles about individuals - where, data about movement, usage, etc., of a subscriber is collected over a period of time and linked with his/her IMSI. Such profiling may expose an individual to various kinds of unanticipated risks and above all will deprive an individual of his privacy. Thus, with more and more people accessing voice, Internet, rich communication services, value added services, mobile banking, mobile commerce, etc., through mobile networks, the importance of identity privacy cannot be underestimated.

Current status of identity privacy in global mobility networks

In global mobility networks, three parties are involved, viz., the user equipment that the subscriber carries with him/her (say U), the home network with which the subscriber is registered (say H), and the visited network that allows the subscriber to access mobile services while h/she is roaming outside the home networks service area (say V). U and H shares a long term secret key K_i and a set of one way hash functions (f_1 - f_5). U communicates with V through wireless link, whereas communication between V and H happens through wired medium.

For access security in global mobility systems, an Authentication and Key Agreement (AKA) procedure, which is a challenge response mechanism, is executed between U and H. During this procedure, U and H mutually authenticate each other.

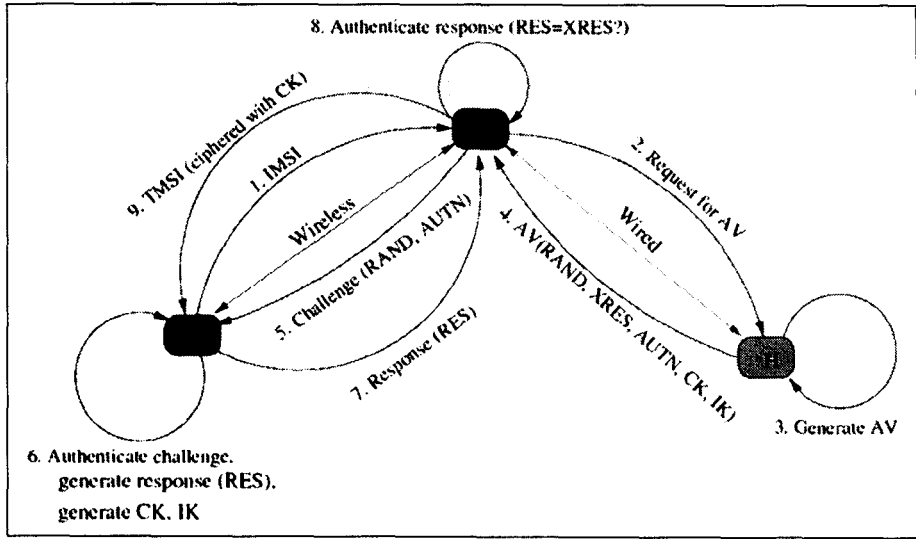


Fig. 4.1: Authentication and Key Agreement Procedure

In Universal Mobile Telecommunication System (UMTS), a commercially successful global mobility network, the AKA procedure (Fig. 4.1) is carried out in the following two stages [8] [9]:

- In the First stage, U presents its IMSI to V. With the help of IMSI, V obtains the security credentials of U in the form of an Authentication Vector (AV) from H. AV are generated using K_i and the hash functions $f1$ - $f5$ (Fig. 4.2). It contains a random number RAND, an authentication token AUTN, an expected response XRES, a cipher key CK and an integrity key IK.
- In the second stage, V utilizes AV to perform mutual authentication with U through a challenge response mechanism. To begin the process, V challenges U by transmitting RAND and AUTN. U in turn, generates response RES, CK, IK and AUTN. For this, $f1$ - $f5$, RAND and K_i are used. It then authenticates V by verifying the generated AUTN against the received AUTN. After this, U sends a response to V's challenge, by transmitting RES. Finally, V authenticates U by comparing RES with XRES. At the successful end of this AKA procedure, a Cipher Key

(CK) and an Integrity Key (IK) are established between U and V. These two keys enable communication over the otherwise vulnerable radio link (between U and V) to happen in a secured and reliable way.

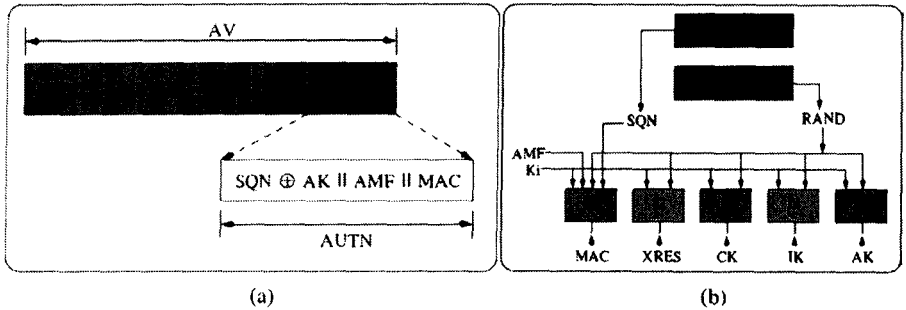


Fig. 4.2: AV Generation

Since, identity presentation during an AKA precedes all other security, in the first stage, U is forced to send its identity (IMSI) in clear text to V through the vulnerable wireless link between them. This makes identity privacy of the subscriber vulnerable to eavesdroppers in wireless link.

In order to provide identity privacy to the subscribers in prominent global mobility systems like UMTS, LTE, etc., the permanent identity (i.e., the IMSI) of the subscriber is replaced by pseudonyms. Instead of the IMSI, short lived pseudonyms are used for identity presentation. Pseudonyms are allotted to U by V. A mapping between a pseudonym and its corresponding IMSI is maintained by V, so that V can resolve it back to corresponding IMSI when required. While generating/allocating pseudonyms, the following is ensured:

- A pseudonym should not have any correlation with any previously generated pseudonym.
- It should not be possible for anyone except V to resolve the corresponding IMSI from a given pseudonym.
- A pseudonym is allotted to U only after a secured channel is established between U and V.

Identity Privacy Vulnerabilities in Global Mobility Networks

In spite of the pseudonym based security mechanism used for identity privacy (discussed in Section III), there are situations when the identity privacy of a subscriber becomes vulnerable. Such situations are as follows:

- U is switched on for the first time and has not yet received a

pseudonym: In such a situation, it is forced to present its identity by transmitting its IMSI in clear-text through the radio link.

- V cannot map a presented pseudonym to its corresponding IMSI: In such a situation (that may arise due to reasons like database failure, etc.), V has a provision to request U for its IMSI. Such a request requires U to transmit its IMSI in clear-text through the radio link.
- A new V cannot contact the old V for the pseudonym-to-IMSI mapping of a roaming subscriber: When a subscriber moves into the region of a new V (say V_n), U will present its identity to V_n through the pseudonym allocated to it by the previous V (say V_o). In order to request for a new set of AV from H, V_n will need to have knowledge of the IMSI. Normally, this will be obtained by presenting the pseudonym to V_o. However, in case V_o cannot be contacted, V_n will be forced to ask U for its IMSI. The later will then have to be transmitted in clear-text over the radio link by U. This vulnerability can in fact be exploited by an attacker who can masquerade as a new V.

Related Work

In the early days of mobile communications, when researchers started recognizing identity privacy as an important security issue, several schemes and protocols were proposed to improve identity privacy of the subscriber over the radio access link between U and V. Some of these schemes and protocols are the ones proposed by Asokan [10], Lin et al. [11], Horn et al. [12], Park et al. [13], Barbeau et al. [7], Juang et al. [14], Forsberg et al. [15], etc. A basic assumption in these proposals is that V is trustworthy.

In present day context, where V can practically be forged or compromised, such trust is difficult. Realizing this, researchers are now concentrating on providing identity privacy to the subscriber over the entire path between U and H. In these solutions, the need to protect the identity privacy of a subscriber from even V is well recognized.

Several schemes and protocols that envisage protecting the identity privacy of the subscriber from V in global mobility networks are proposed. In many of these proposals, asymmetric key cryptography is used, viz., the protocols proposed by Samfat et al. [16], Godor et al. [17][18], Yang et al. [19], Li et al. [20], He et al. [21], Feng et al. [22], etc. In many proposals, a hybrid approach is adopted, where a combination of both asymmetric key and symmetric key crypto systems are used. Some of these protocols are the ones proposed by, He et al. [23], Varadharajan et al. [24], Al-Fayoumi et al. [25], Zhu et al. [26], Koien et al. [27], etc. Recently, many schemes were proposed, where computationally light techniques that use symmetric key based crypto systems, hash functions,

XOR operations, temporary identities, alias, etc., are used. Some of these protocols are the ones proposed by Sattarzadeh et al. [28], Tang et al. [29], Pereniguez et al. [30], He et al. [31], Zhou et al. [32], He et al. [33], Chen et al. [34], Liu et al. [35], Jiang et al. [36], etc.

Roaming agreements simplified

Recent research conducted to enhance identity privacy in global mobility networks, as discussed in Section V, envisage protecting the subscribers' identity from visited networks (V) which may even be forged or compromised. In this, the author of this paper perceives a supplementary benefit, which is that of simplification in the roaming agreement procedure between mobile operators.

Roaming agreements/pacts with third party operators to provide service in a location where an operator has not set up its own infrastructure is a common practice. Such agreements allow a subscriber of one operator to use the access service of another operator when inside the latter's coverage area.

The current status of identity privacy in global mobility networks requires establishing the following trust relationships with respect to its subscribers' identity, i.e., the IMSI (Fig. 4.3).

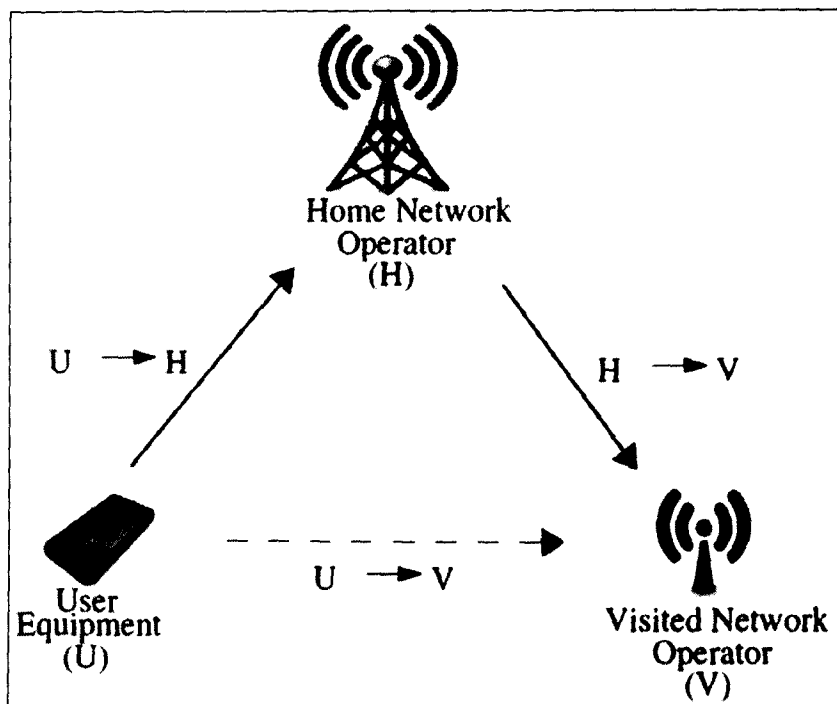


Fig. 4.3: Current Trust Model

1. **U→H:** As the **U** is registered with **H**, it trusts **H** with its IMSI.
2. **H→V:** **H** confers full trust in **V** with regards to the IMSI of the subscriber.
3. **U→V:** This trust relation is a transitive outcome of the previous two trust relations, because of which, the **U** also has to trust **V** with its IMSI.

Here, the first trust relationship requirement is unavoidable. However, the other two trust relationships requirements are due to the operator's obligation towards the subscribers to ensure adequate identity privacy while they are roaming within the coverage area of another operator. To establish such trust relationships, it becomes necessary for the home network operator to have elaborate negotiations or agreements with the visited network operator. Such negotiations, limits the ease and span of extending services beyond an operator's own circle/zone. A visited network operator, with which there is no prior roaming agreement, cannot be trusted by the home network operator when it comes to its subscribers' identity privacy.

With the increase in demand for any-time any-where service, there is need for a paradigm shift, such that the requirement of trust on the visited network operator is relaxed. Recent research in the field of identity privacy that envisages protecting the subscribers' identity even from **V** fulfils this requirement. With this new approach, the need to trust **V** with the identity of the subscriber is eliminated. Thus, the only trust relationship that an operator is expected to take care of with respect to the subscriber's identity is the following (Fig. 4.4).

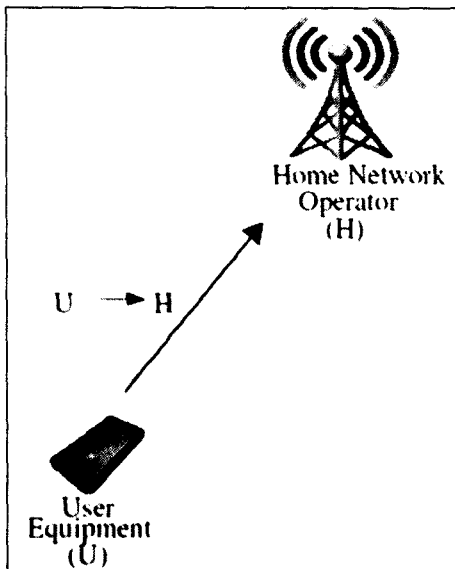


Fig. 4.4: Simplified trust model.

2. **U→H:** As the **U** is registered with **H**, it trusts **H** with its IMSI.

Therefore, this new approach to enhance identity privacy of the subscriber in global mobility networks would be a step towards enabling the home network operator to have simplified negotiations/agreements with the other network operators.

Conclusion

In today's context, when mobile operators strive to provide wider coverage, roaming agreements/pacts with third party operators to provide service in a location where an operator has not set up its own infrastructure is a common practice. Roaming allows a subscriber of one operator to use the access service of another operator when inside the latter's coverage area. Recent research to enhance identity privacy in global mobility networks envisages protecting the identity of the subscriber even from the visited network operator. If this becomes a reality in practice, it would remove the need for the home network operator to trust the visited network operator with the subscriber's identity. This relaxation is perceived to simplify roaming agreements. In this paper, an effort was made to elaborate on this possibility.

References

1. Whalen, T. Mobile devices and location privacy: Where do we go from here? *IEEE Security & Privacy* 9(6), 61-62, 2011.
2. TWB. Mobile phone access reaches three quarters of planet's population, The World Bank Press Release. <http://www.worldbank.org/en/news/2012/07/17/mobile-phone-access-reaches-three-quarters-planets-population>, 2012.
3. 3GPP. Numbering, addressing and identification, Technical Specification. <http://www.3gpp.org/ftp/Specs/html-info/23003.htm>, 2012.
4. Liu, Z. Y. et al. A fast suffix matching method in network processor, in *IEEE International Conference on Computational Intelligence and Security (CIS' 08)*, Suzhou, China, 405-410.
5. Barbeau, M. & Robert, J. M. Perfect identity concealment in UMTS over radio access links, in *IEEE International Conference on Wireless And Mobile Computing, Networking And Communications (WiMob 2005)*, Montreal, Canada, 72-77.
6. Boman, K. et al. UMTS security, *Electronics & Communication Engineering Journal* 14(5), 191-204, 2002.
7. Niemi, V., Nyberg, K. & Wiley, J. *UMTS Security*, Wiley, United States, 2003.
8. Koien, G. M. An introduction to access security in UMTS, *IEEE Wireless Communications* 11(1), 8-18, 2004.
9. Xenakis, C. & Merakos, L. Security in third generation mobile networks, *Computer communications* 27(7), 638-650, 2004.

10. Asokan, N. Anonymity in a mobile computing environment, in IEEE Workshop on Mobile Computing Systems and Applications (WMCSA 1994), Santa Cruz, CA, 200-204.
11. Lin, H. Y. & Harn, L. Authentication protocols for personal communication systems, ACM SIGCOMM Computer Communication Review 25(4), 256-261, 1995.
12. Horn, G. & Preneel, B. Authentication and payment in future mobile systems, in European Symposium on Research in Computer Security (ESORICS' 98), Louvain-la-Neuve, Belgium, 277-293.
13. Park, J. et al. Wireless authentication protocol preserving user anonymity, in Symposium on Cryptography and Information Security (SCIS 2001), Oiso, Japan, 159-164.
14. Juang, W. S. & Wu, J. L. Efficient 3GPP authentication and key agreement with robust user privacy protection, in IEEE Wireless Communications and Networking Conference (WCNC 2007), Kowloon, Hong Kong, 2720-2725.
15. Forsberg, D. et al. Enhancing security and privacy in 3GPP E-UTRAN radio interface, in IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2007), Athens, Greece, 1-5.
16. Samfat, D. et al. Anonymity and untraceability in mobile networks, in ACM International Conference on Mobile Computing and Networking (MobiCom' 95), Berkeley, California, USA, 26-36.
17. Godor, G. et al. Novel authentication algorithm of future networks, in IEEE International Conference on Networking, International Conference on Systems and International Conference on Mobile Communications and Learning Technologies (ICNICONSMCL' 06), Mauritius, 80-80.
18. Godor, G. & Imre, S. Novel authentication algorithm - public key based cryptography in mobile phone systems, International Journal of Computer Science and Network Security 6(2B), 126-134, 2006.
19. Yang, G. et al. Anonymous and authenticated key exchange for roaming networks, IEEE Transactions on Wireless Communications 6(9), 3461-3472, 2007.
20. Li, X. & Wang, Y. Security enhanced authentication and key agreement protocol for LTE/SAE network, in IEEE International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM'11), Wuhan, China, 1-4.
21. He, D. et al. Privacy-preserving universal authentication protocol for wireless communications, IEEE Transactions on Wireless Communications 10(2), 431-436, 2011.
22. Feng, T. et al. Anonymous identity authentication scheme in wireless roaming communication, in IEEE International Conference on Computing Technology and Information Management (ICCM 2012), Berlin, Germany, 124-129.
23. He, Q. et al. The quest for personal control over mobile location privacy, IEEE Communications Magazine 42(5), 130-136, 2004.
24. Varadharajan, V. & Mu, Y. Preserving privacy in mobile communications: A hybrid method, in IEEE International Conference on Personal Wireless Communications (ICPWC 1997), Mumbai, India, 532-536.

25. Al-Fayoumi, M. et al. A new hybrid approach of symmetric/asymmetric authentication protocol for future mobile networks, in *IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMOB 2007)*, New York, USA, 29-29.
26. Zhu, J. & Ma, J. A new authentication scheme with anonymity for wireless environments, *IEEE Transactions on Consumer Electronics* 50(1), 231-235, 2004.
27. Koien, G. & Oleshchuk, V. Location privacy for cellular systems; Analysis and solution, in *International Workshop on Privacy Enhancing Technologies (PET 2005)*, Cavtat, Croatia, 40-58.
28. Sattarzadeh, B. et al. Improved user identity confidentiality for UMTS mobile networks, in *IEEE European Conference on Universal Multiservice Networks (ECUMN '07)*, Toulouse, France, 401-409.
29. Tang, C. & Wu, D. O. Mobile privacy in wireless networks-revisited, *IEEE Transactions on Wireless Communications* 7(3), 1035-1042, 2008.
30. Pereniguez, F. et al. Privacy-enhanced fast re-authentication for EAP-based next generation network, *Computer Communications* 33(14), 1682-1694, 2010.
31. He, D. et al. A strong user authentication scheme with smart cards for wireless communications, *Computer Communications* 34(3), 367-374, 2011.
32. Zhou, T. & Xu, J. Provable secure authentication protocol with anonymity for roaming service in global mobility networks, *Computer Networks* 55(1), 205-213, 2011.
33. He, D. et al. Design and validation of an efficient authentication scheme with anonymity for roaming service in global mobility networks, *Wireless Personal Communications* 61(2), 465-476, 2011.
34. Chen, C. et al. Lightweight and provably secure user authentication with anonymity for the global mobility network, *International Journal of Communication Systems* 24(3), 347-362, 2011.
35. Liu, H. & Liang, M. Privacy preserving registration protocol for mobile network, *International Journal of Communication Systems*. <http://onlinelibrary.wiley.com/doi/10.1002/dac.2426/full>, 2012.
36. Jiang, Q. et al. An enhanced authentication scheme with privacy preservation for roaming service in global mobility networks, *Wireless Personal Communications* 68(4), 1-15, 20.

Wireless Sensor Network Performance Analysis through Ad-Hoc Routing Protocols using Qualnet

*Hrituparna Paul
Prodipto Das*

Abstract

This paper investigates & undertakes simulation based study of Ad- hoc Routing Protocols in Wireless Sensor network. Here comparison of four Routing Protocols AODV, DYMO, DSR and FISHEY is done in WSN Networks. In this paper we shall study different performance parameters in a Random Waypoint Mobility Model by varying the number of node sandal so changing the maximum speed of a node, such as Average Throughput, Average End to End Delay, Average Jitter, Average PDR(Packet Delivery Ratio) and Total Packets Received. Analysis and performance study is done using QualNet 5.0 Simulator.

Keywords: *Wireless sensor networks, Routing Protocols, Jitter, End to end delay, PDR, Throughput*

Introduction

Sensor networks are the key to assemble the Information needed by smart environments, whether in buildings, utilities, industrial, home, shipboard, transportation systems automation, or elsewhere and provide a wireless communication infrastructure, observe and respond to phenomena in the natural environment and in our physical and cyber infrastructure [1]. In Ad hoc networks [2] mobile nodes that are self organizing connected by wireless links and are free

to move independently and randomly and has the capability to change its links frequently and infrastructure less networks that can be constructed temporarily. A routing protocol specifies how routers communicate with one another, broadcast information that allows them to select their path between any two given nodes on an ad-hoc network. It distributes this information first among its neighbours, and then to the whole of the network. In this way, the routers achieve knowledge of the topology of the network. Some examples of routing protocols available for Ad-hoc networks are AODV, CGSR, DSDV, DSR, OLSR, WRP, ZRP, AOMDV, FISHEY, etc.[3]. Mobility models represent the movement of nodes and how their location, velocity and acceleration change with respect to time. In the study of a new Mobile ad hoc network protocol, it is important to simulate the protocol and evaluate its protocol performance.

The rest of the paper is organized as follows: section- 2 describes the RWM model, Section 3 which gives a brief description and overview of routing protocols AODV, DSR, DYMO and FISHEY. Section 4 explained the Simulation Result and conclusion are described in section 5.

Random Waypoint Mobility Model

The random waypoint model is a commonly used mobility model for simulations of wireless Communication networks and for the movement of mobile users, and how their location, velocity and acceleration change overtime. Mobility models are used for simulation purposes when new network protocols are evaluated. The Randomway point model was first proposed by Johnson and Maltz[4]. It is one of the most popular mobility models and the "benchmark" mobility model to evaluate other Mobile ad hoc network (MANET) routing protocols, because of its simplicity and wide availability.

Random Way point (RWP) model is a commonly used synthetic model for mobility, e.g., in Ad-Hoc networks. It is an elementary model which describes the movement pattern of independent nodes by simple terms. This model includes:

- Each node moves along a zigzag line from one way point P_i to the P_{i+1}
- Optionally, the nodes may have so-called "thinking times" when they reach each way point before continuing on the next leg, where durations are independent and identically distributed random.

To be more specific, the destination, speed and direction are all chosen randomly and independently of other nodes. This type of model has been used in many simulation studies. In random waypoint mobility model, the nodes can randomly select a position, move towards it in a straight line at a constant speed that is randomly selected from a range, and pause at that destination. The node repeats this, throughout the simulation [5].

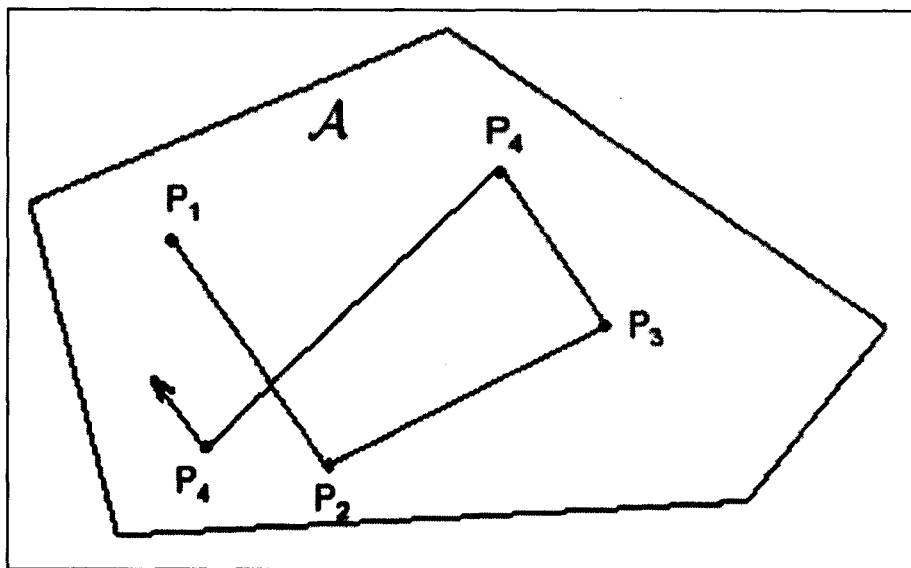


Fig. 5.1: Random Waypoint Model

Brief Description and Overview of Routing Protocol

Ad-hoc On Demand distance Vector routing protocol (AODV)

AODV [6,7] is a reactive routing which [The AODV Routing protocol] [7] uses an on-demand approach for finding routes, that is, a route is established only when it is requisite by a source node for transmitting data packets. AODV uses sequence numbers to make certain the freshness of routes. It is loop-free, self-starting, and scales to large records of mobile nodes. The AODV protocol uses route request (RREQ) messages flooded through the set of connections in order to discover the paths required by a source node. AODV allows mobile nodes to locate out routes quickly for new destinations, and does not necessitate nodes to maintain routes to destinations that are not in dynamic communication. It allows nodes to act in response to link breakages and a alteration in network topology in a timely manner and the operation [8] of AODV is loop-free. When a route to a new destination is necessary, the source broadcasts a RREQ message to find a route to the requisite destination. A transitional node that receives a RREQ replies to it using a route reply message only if it has a route to the destination whose analogous destination sequence number is greater or equal to the one contained in the RREQ. Another important point to note the RREQ also contains the most recent sequence number for the destination of which the source node is responsive. A node receiving the RREQ may send a route reply (RREP) if it is either the destination or if it has a route to the destination with equivalent

sequence number greater than or equal to that contained in the RREQ. If this is the case, it unicast a RREP reverse to the source. Otherwise, it rebroadcasts the RREQ. Nodes maintain track of the RREQ's source IP address and broadcast ID and if they take delivery of a RREQ which they have already processed, they discard the RREQ and do not to the fore it.

Dynamic Source Routing (DSR)

Dynamic Source Routing (DSR) is a routing protocol for wireless mesh networks and is based on a technique known as source routing. DSR allows the network to be completely self-organizing and self-configuring, devoid of the need for any pre-existing network infrastructure.

The Dynamic Source Routing protocol [9] is composed of two main mechanisms route discovery and route maintenance. In the Route Discovery mechanism a source node wishing to drive a packet to a destination node, ascertains a source route to the destination. In Route Maintenance mechanism a node wishing to send a packet to a destination is able to perceive, while using a source route to the destination, if the network topology has altered such that it can no longer make use of its route to destination because a link along the route no longer works. And in case when Route Maintenance indicates a source route is broken, source can attempt to bring into play any other route, it happens to know to destination, or it can invoke Route discovery again to find a new route for subsequent packets to destination.

Dynamic MANET On-demand routing protocol (DYMO)

DYMO [10] handles an extensive variety of mobility patterns by dynamically determining routes on-demand. It also handles a wide selection of traffic patterns. The fundamental operations of the DYMO routing protocol are route discovery and route maintenance.

During route discovery, a DYMO router initiates flooding of a Route Request message (RREQ) throughout the network to come across a route to a particular destination, via the DYMO router accountable for this destination. During this hop-by-hop flooding process, each transitional DYMO router receiving the RREQ message records a route to the originator. When the target's DYMO router receives the RREQ, it records a route to the originator and responds with a Route Reply (RREP) unicast hop-by-hop in the direction of the originating DYMO router. Each intermediate DYMO router that receives the RREP creates a route to the target, and then the RREP is unicast hop-by-hop on the way to the originator. When the originator's DYMO router receives the RREP, routes have been established between the originating DYMO router and the target DYMO router in both directions.

Route maintenance consists of two operations. In order to conserve routes

in use, DYMO routers expand route lifetimes upon successfully forwarding a packet. In order to act in response to changes in the network topology, DYMO routers keep an eye on traffic being forwarded. When a data packet is received for forwarding and a route for the destination is not known or the route is broken down, then the DYMO router of the source of the packet is notified. A Route Error (RERR) is transmitted to point out the route to one or more affected destination addresses is Broken or misplaced. When the source's DYMO router receives the RERR, it marks the route as broken. Before the DYMO router can forward a packet to the same destination, it has to carry out route discovery again for that destination.

Fisheye State Routing Protocol (FSR)

Fisheye State Routing (FSR) [11] is a Multilevel with Scope technique, table driven routing protocol for ad-hoc network. It is designed to reduce routing overheads in large and fast dynamic changing network. Fish eye normally observers and focus with high detail on the object very close to its focal point. When the object distance increases from the focal point the detail decreases. The same principle is used in Fisheye State routing. FSR maintain topology map a teach node. FSR will not flood or broad cast to evaluate the route. Instead, nodes maintain a link state table based on updated information from the neighbor. A full topology map will be stored in each node of the network. The topological map will be utilized to route discover and route maintenance.

The following are the advantages of FSR.

- Simplicity
- Usage of up-to-date shortest routes
- Robustness to host mobility
- Exchange Partial Routing Update with neighbors

Performance Matrices

Throughput

In a WSN, throughput is measured in terms of successful delivery of data packet within the Threshold time. The data may use different routes and passes across multiple intermediate nodes to reach the destination. Throughput is measured using number of bits of packet received per unit time. Normally throughput is measured as bits per sec.

The following are major factors affecting throughput:

- Packet loss due to network congestion
- Available bandwidth
- Number of Users in the Network

- Data loss due to bit errors
- Improper queuing techniques used
- Slow Start and multiple decrease techniques

Average End-to-End Delay

Average End-to-End delay is a metrics used to measure the performance with time take by a pack to travel across a network from a source node to the destination node. In WSN, sensor nodes switch between an active (on) and a sleeping (off) mode, to save energy. Such Scenario pays a greater latency in the sensor network. Each sensor node with sensed data has to wait for the neighbor sensor node to turn it to active mode from sleep mode [10]. End to end delay evaluates latency when data send by sensor nodes and received by destination node. An end to end delay includes all possible delay caused during route discovery, retransmission delay,

Queuing delay and relay time.

$$D_{\text{end-end}} = N (D_{\text{trans}} + D_{\text{prop}} + D_{\text{proc}})$$

Where,

$D_{\text{end-end}}$ = End-to-End Delay, D_{trans} = Transmission Delay,
 D_{prop} = PropagationDelay, D_{proc} = ProcessingDelay.

Average Jitter

Average jitter is a performance characteristics used to measure deviation from true periodicity eventually of inactivity in packet across a specific network. When a network is stabilized with constant latency will have no jitter. Packet jitter is expressed as an average of the deviation from the network mean latency. Due to data congestion or route changes can cause jitter. In a Wireless sensor networks, multiple sensor nodes may sense the information and forward them to the sink in continuous manner. Due to bottle neck problem or network congestion in the receiver end a delay may occur. This delay causes a deviation from the jitter. Average Jitter in a network increases indefinitely due to improper queuing techniques or configuration errors.

Simulation Result & Discussion

The simulator is used to record the performance parameters is Qual net 5.0.2 developed by SCALABLE Network Technologies. In the Architecture mode of the simulator the scenario is designed in an area of 360000 square meters. Initially if no changes are made to the area then automatically the simulator takes an area of 1500X1500 square meters. Number of Nodes are increased from 5 to 20 in multiples of 5. Network traffic type is chosen as CBR (ConstantBit Rate).

Table 5.1: Scenario Description

Parameters	Values
Simulator	QualNet
Protocols studied	AODV,DSR,DYMO,FI
Number of nodes	20,40
Simulation area	1500*1500
Node	Random waypoint
Movement model	Mobility
Traffic types	CBR sources

Below are the screenshots from the Qualnet simulator when the above mentioned scenario is designed in the architect mode and before the results are analysed through the analyser, the Qualnet software runs the experiment. Several stages are shown below while the simulation takes place, in3 dimensional and X-Y axis view. Each stage shows the performance and movement of nodes randomly, due to random waypoint mobility, broadcasting, data traffic through CBR, forming an ad hoc network and choosing the paths as per the protocols used.

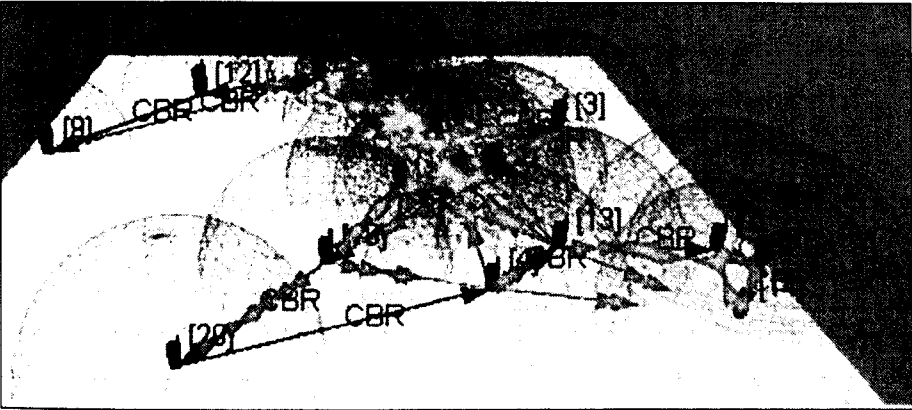


Fig. 5.2: D view of scenario simulation in Qualnet

From Fig. 5.2, it can be observed that DYMO has good throughput at a higher mobility. DSR has a better throughput when compared to FISHEY and AODV for each set of connections for 20 nodes. But with increasing the number of nodes, throughput of FISHEY is better compared to AODV, DSR and DYMO for each set of connections for 40 nodes. FISHEY has good average throughput.

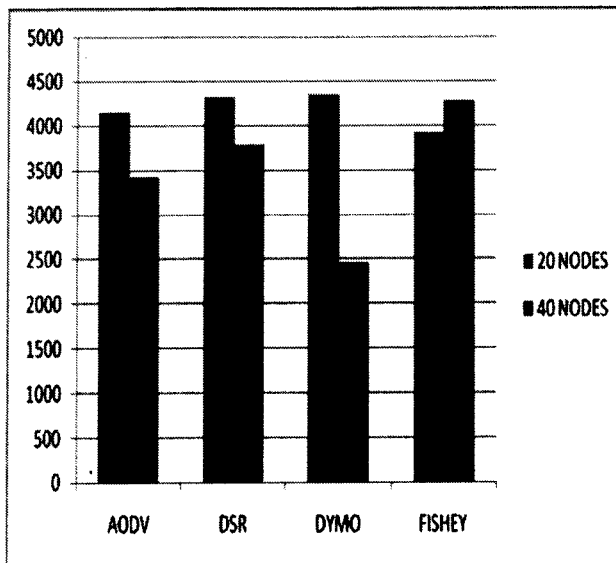


Fig. 5.3: Throughput

Fig.5. 3 shows that FISHEY exhibited the lowest average end-to-end delay for both 20 and 40 nodes, while AODV had the highest delay in case of 20nodes and DSR had the highest delay in case of 40 nodes.

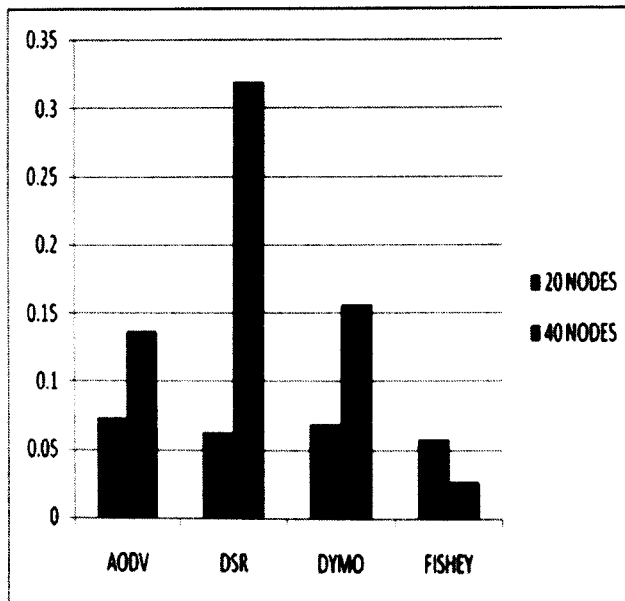


Fig. 5.4: Average End to End Delay(s)

From Fig. 5.4, it can be observed that FISHEY has lowest Jitter in case of 20 nodes. For 40 nodes DYMO has lowest jitter. But the Average Jitter of FISHEY is lowest.

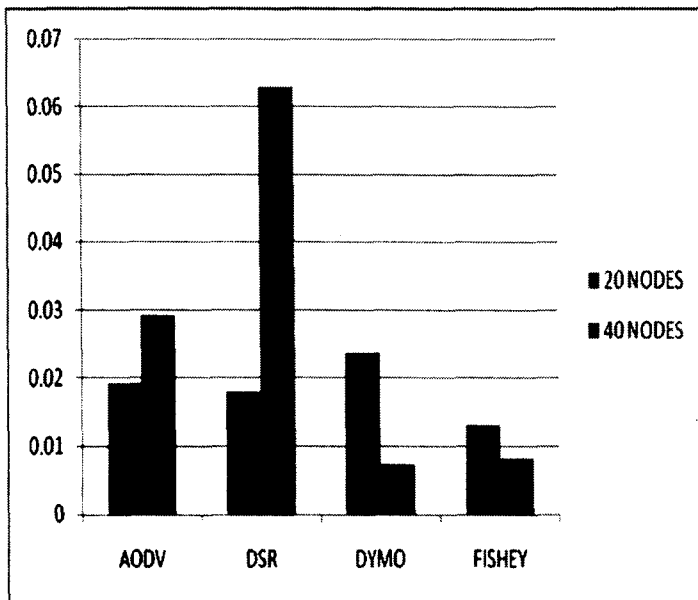


Fig. 5.5: AverageJitter(s)

Conclusion

In this paper different routing protocols such as AODV, DSR, DYMO and FISHEY is evaluated. With the help of simulation we compared these protocols with different number of nodes. We measure the Average Jitter, Through put and End to end delay as performance matrices. On the basis of simulation result as a conclusion we can say that overall performance of FISHEY is better for different number of nodes.

References

1. V. Thapar, B. Jain and V. Sahni, "Routing protocols using random way point mobility model in wireless sensor networks" *International Journal on Computer Science and Engineering* (ISSN: 0975-3397) Vol. 3No.8 August2011 page 3059.
2. Ritika & Nipur, "Performance Evaluation of Reactive, Proactive and Hybrid Routing Protocols Based on Network Size for MANET", *International Journal of Computer Science and Security (IJCSS)*, Volume (6): Issue (1), pp. 34-42, 2012.
3. A. Goyal, S. Vijay and D. K. Jhariya "Simulation and Performance Analysis of Routing Protocols in Wireless Sensor Network using Qualnet" *International Journal of Computer Application* (0975-8887) Volume 52-No.2, August2012.

4. Johnson, D. B.; Maltz, D.A. (1996). "Dynamic Source Routing in Ad Hoc Wireless Networks". Mobile Computing. The Kluwer International Series in Engineering and Computer Science 353 .p.153.doi:10.1007/978-0-585-29603-6_5. ISBN978-0-7923-9697-0.
5. Bisht, R. Mishra, R. Dhillon "Mobility Based Performance Analysis of Wireless Sensor Networks", International Journal of Computer, Electronics & Electrical Engineering (ISSN: 2249-9997) Volume 2-Issue 2.
6. C. E. Perkins and E. M. Royer, "Ad-hoc On-Demand Distance Vector Routing", Proceeding of 2nd IEEE Workshop, Mobile Computing System Applications, pp:90-100, 1999.
7. Perkins, E. Belding-Royer, and S. Das, "Ad hoc On-Demand Distance Vector (AODV) Routing, "draft-ietf-manet-aodv-13.txt, Feb. 2003.
8. S.A.Adel & P.A.Tijare2; "Performance Comparison of AODV, DSDV, OLSR and DSR Routing Protocols in Mobile Ad Hoc Networks", International Journal of Information Technology and Knowledge Management July-December 2010, Volume 2, No. 2, pp. 545-548.
9. J. Broch, D. Jhonson, and D. Maltz, "The dynamic source routing protocol for mobile ad hoc networks for IPv4" IETF RFC 4728, Feb 2007.
10. I. Chakeres and C. Perkins, "Dynamic MANET On-Demand (DYMO) Routing," IETF Internet-Draft, draft-ietf-manet-dymo-23, Oct. 2012.
11. G. Pei, M. Gerla, and T. W. Chen, "Fisheye State Routing in Mobile Ad-Hoc Networks, "In Proceedings of the 2000 ICDCS workshops, Taipei, Taiwan, Apr. 2000.

Neural Network Based Modified AODV Routing Protocol in VANET

Soumen Saha

Soumen Roy

Utpal Roy

Devadutta Sinha

Abstract

The latest mobile communications is two types of ad hoc networks are introduced; one is Mobile Ad Hoc Network (MANET) and second one is Vehicular Ad Hoc Network (VANET). VANET is based on IEEE 802.11b wireless standard. This helps to communicate vehicle to vehicle and vehicle to roadside communications. According to Federal Communications Commission (FCC) suggests for VANET frequency spectrum of 75 MHz in the range of 5.850 GHz to 5.925 GHz. It communicate from one vehicle (source) to another vehicle (destination) through different vehicles (intermediate nodes). A numbers of different routing protocols for communication, ie multimedia data, text data etc. from one vehicle (node) to another vehicle are existing. The Ad hoc On-Demand Distance Vector (AODV) routing algorithm is one of the popular routing protocols for ad-hoc mobile networks. AODV is used for both unicast and multicast routing. In this paper, we propose and implement in the NCTUns-6.0 simulator neural network based Modified AODV routing protocol considering Power, TTL, Node distance and Payload parameter to find the optimal route from the source station (vehicle) to the destination station in VANET communications.

Keywords: *AODV, Neural Network, NCTUns-6.0, VANET, Routing Protocol.*

Introduction

Vehicular Ad-hoc Networks (VANETs) [1]-[2] identifies an emerging technology, particularly challenging a class of Mobile Ad Hoc Networks (MANETs). VANETs are distributed; self-organizing communication networks built up by moving vehicles, and is thus characterized by very high node mobility with limited degrees of freedom in the mobility patterns. Hence, ad hoc routing protocols must adapt continuously to these unreliable and unethical conditions; in the event of growing effort in the development of communication protocols which are originated to vehicular networks. One of the critical aspects when evaluating routing protocols for VANETs is the employment of mobility models that reflect as closely as possible to the real behavior of vehicular traffic. This notwithstanding, using simple random-pattern, the graphs constrained mobility models is a common practice among researchers working on VANETs. There is no need to say that such models cannot describe vehicular mobility in a realistic way, since they ignore the peculiar aspects of vehicular traffic, such as cars acceleration and retardation in presence of nearby vehicles, queuing at roads intersections, traffic bursts caused by traffic lights, and traffic congestion or traffic jams. All these situations greatly affect the network performance, since they act on network connectivity, and this makes vehicular specific fundamental performance evaluations at the time of studying routing protocols for VANETs.

Wireless technologies [3]-[4] are extended to ad hoc networks like Mobile Ad Hoc Network (MANET) and Vehicular Ad Hoc Network (VANET) [5]-[6], [7]-[8]. Ad hoc networks are one type of network that offers communications within a certain range of areas; even connect to wide areas via basic mobile network and Internet. One of the authors has already published his useful and important findings of various routing protocols [5], mainly many variants AODV applicable in MANET. This study is the modest approach towards the justification for application of AODV routing protocol of MANET in Vehicular Transmission [6]-[7].

VANET is one special type or a subset of MANET, which is exhaustively communicating in a group of moving vehicles or extending to a wide zone through basic mobile infrastructures or services and internet facility. Mobile ad hoc networks, also known as short-lived networks, are autonomous systems of mobile nodes forming network either in the absence of any centralized support or presence of a basic network. The basic mobile communications is a costlier one and cannot afford communications covering all regions in the world, especially in remote and less dense populated area, wide sea or desert, etc.

In the areas where the normal mobile communications are not economically viable, ad hoc network is the best solution. MANET and VANET are self forming networks, i.e., they can work without any centralized control like Base Station

(BTS) or Switch (BSC, MSC) in mobile network or Access Point (AP) in local area network (LAN) [1]-[6]. BTS is a Base Trans-receiver Station from where air interface link is connected to mobile subscriber or instrument (MS). BSC is a Base Switching Center and MSC is a Main or Mobile Switching Center. A mobile path is connected from BTS to MSC via BSC or vice versa through optical fiber or microwave link and call connection is controlled by MSC only. Each terminal or node (either mobile phone or computer) in a MANET or a VANET acts both as a data or voice terminal and router or switch. A node in a cell communicates with other nodes in its transmitting range through wireless medium only. Thus a VANET is a small structure or subset of a MANET. In a VANET, certain number of moving vehicles in a small region constitutes a cell. It means that the range of wireless signal, i.e., transmitting zone from a moving vehicle is within a limited area. A vehicle, called a node, can do transmitting, receiving, and routing (connecting) to other nodes without help of any switch like Base Station (BTS) in basic mobile network or Access Point (AP) in LAN. Also the moving vehicle in a VANET cell can be connected to other nodes in another cell or other network like basic mobile network, Internet, etc. Thus total connectivity in a VANET is assured. VANETs are also known under different name like Dedicated Short Range Communications (DSRC), Inter Vehicle Communications (IVC), etc. Number of projects have been launched for VANET, e.g., FleaNet in USA, FleetNet in Germany, ITS in Japan, etc. [1]-[7], [11].

The motivation of a VANET project is to create a new algorithm or protocol or modify the existing one for use in vehicular environment. Thus VANET helps the drivers of vehicles to communicate the information in form of voice, data, image, multimedia, etc. Also it ensures safe journey by minimizing road accidents, diverting or instructing the vehicle's direction in less populated roads avoiding traffic jam, etc. Vehicles in a VANET are having high degree of mobility, i.e., the vehicles are moving very fast, especially in high ways. As a result the two vehicles are in a direct communication range staying about one minute time only, i.e., two vehicles remain in one cell about one minute time when they are moving parallel direction or even less than one minute when they are in opposite direction [2]-[4]. For this, VANET cell configuration and number of nodes present in a particular cell with applicable routing technique is changing in nature.

VANET is based on IEEE 802.11 wireless standard which is mainly framed for WLAN, WiFi, MANET, and VANET. Demerits of IEEE 802.11 standards are that no retransmission is possible for failed broadcast transmissions and that the contention window (CW) size fails to change because of the lack of MAC (Medium Access Control) level recovery. Therefore, CW is held constant for broadcast transmission. It is noteworthy mentioned that the probability of reception of a broadcast message decreases as the distance from the sender increases and under saturated conditions the probability of reception of a message is very low.

An advanced artificial intelligence (AI) based Modified Ad hoc On-Demand Distance Vector (aiAODV) routing protocol applying fuzzy neural network algorithm is proposed in this paper. Generally the scientists may not be able to provide error free data or knowledge using fuzzy logic system. For that a neuro fuzzy system can be used to tune the system and reject unnecessary or redundant fuzzy rules. A neuro fuzzy system has multilayers that embed the fuzzy system. By applying this fuzzy neural network in aiAODV routing protocol we are able to determine the optimum route (path) from the source vehicle to the destination.

Architecture of MANET and VANET

Initially IEEE 802.11 is implemented on WLAN at a speed of 1 or 2 Mbps (very slow) in 1997 A.D. Then IEEE 802.11 protocol family is upgraded to different versions. 802.11a uses Orthogonal Frequency Division Multiplexing (OFDM) modulation to deliver up to 54 Mbps in the wider frequency 5 GHz ISM band. IEEE 802.11b applies High Rate Direct Sequence Spread Spectrum (HR-DSSS) to achieve 11 Mbps in 2.4 GHz ISM band. IEEE 802.11g implements OFDM modulation, but operates narrow 2.4 GHz ISM band. The Federal Communications Commission (FCC) suggests for VANET frequency spectrum (bandwidth) of 75 MHz in the range of 5.850 GHz to 5.925 GHz in USA. Seven channels are fragmented, having each 10 MHz bandwidth. Out of these, six channels are used for services and one channel is used for control purpose like broadcast services, e.g., safety message, announcement, etc.

IEEE 802.11 protocols apply Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) with acknowledgements for reliable communications and avoiding collision among packets. In a VANET, Medium Access Control (MAC) layer determines a contention based access protocol, termed Distributed Coordination Function (DCF). Actually, MAC sublayer determines how the channel is allocated, i.e., whose turn to transmit next. Above MAC, there is Logic Link Control (LLC) layer to hide the difference amongst different IEEE 802 variants. The main drawback of the DCF is hidden terminal problem, hence Quality of Service (QoS) is not guaranteed in wireless network like VANET.

If transmitting power of a node (vehicle) increases in a VANET, signal from the node spreads more area, i.e., the cell size becomes large, but throughput of the network, i.e., information handling capacity will decrease. One hop count is the minimum distance between one node to another neighborhood node. Again if the number of hops (nodes or cells) increases in a region, expected path of life decreases. The transmission signal in a VANET is interference due to multipath fading, different type of noises, power supply fluctuation and out of cover range (mobility), etc. Therefore maximum care has to be taken to retrieve the signal in original form by eliminating noise using multiple antennas called diversity antennas, low noise amplifier (LNA) filter in the receiver circuit of the VANET node or

vehicle. Also higher bandwidth (75 MHz) permits to use spread spectrum technique in VANET communications by adding number of extra bits in a frame or packet which ensures error free communication in a real time basis. Therefore, VANET communication becomes more reliable and effective one for short range of ad hoc networks using through vehicles.

Routing Protocols in Ad Hoc Network

In this section, we discuss the routing protocols available for ad hoc networks like MANETs and VANETs. Routing determines process of communicating information from one station to another station of the network through passing one or more intermediate stations, called intermediate nodes. An ad hoc routing protocol [3]-[11] is a procedure or standard, that controls how vehicles (nodes) decide which way to direct packets called traffic among nodes (computing devices). Routing protocols of mobile ad hoc network require different approaches from existing internet protocols (IPv4 or IPv6), since most of the existing internet protocols are designed to support routing in a network with fixed structure. Lots of various routing protocols are coming to use. Proposed routing solutions are classified into seven types: proactive or table-driven, reactive or on-demand, hybrid, hierarchical, geographical, power-aware, geographical multicast protocols [4]-[11]. Out of these routing protocols, Ad Hoc On-Demand Distance Vector (AODV) routing protocol is an effective one.

Reactive or On-Demand Routing Protocol

This type of protocols finds a route on demand by flooding the network with route request packets. Actually, a route is decided on the availability of least distance, less overload or overhead, less consumption of electrical power, traffic solution, etc. and this protocol is changing in nature. It initiates a route discovery process, which goes from one node to the other until it reaches to the destination or an intermediate node which has a route to the destination. The main disadvantages of such algorithms are as followings:

- i. High latency time in route finding.
- ii. Excessive flooding can lead to network clogging, i.e., the network is blocked or congested.
- iii. It is the responsibility of the route request receiver node to reply back to the source node about the possible route to the destination. The source node uses this route for data transmission to the destination node.

Some of the better known on-demand protocols are such as Robust Secure Routing Protocol (RSRP), Multirate Ad Hoc On-demand Distance Vector (M-AODV) Routing Protocol, Modified AODV Routing Protocol Explored by Swarm Intelligence, Reliable Ad Hoc On-demand Distance Vector (R-AODV) Routing

Protocol, Minimum Exposed Path to the Attack (MEPA) in Mobile Ad Hoc Network (MANET), Ant-based Routing Algorithm for Mobile Ad Hoc Networks, Admission Control enabled On-demand Routing (ACOR), Dynamic Source Routing (DSR), and Temporary Ordered Routing Algorithm (TORA), etc. Among these protocols, AODV routing protocol is more useful in ad hoc mobile networks. We are discussing the AODV protocol in detail.

Features of AODV Routing Protocol

Ad hoc On-Demand Distance Vector (AODV) routing protocol [7]-[11] is more popular and effective one in ad hoc networks like MANET and VANET communications. It is jointly developed in Nokia Research Center, University of California, Santa Barbara and University of Cincinnati. Since AODV is a reactive protocol, it establishes a route to a destination only on demand. It is capable of both unicast and multicast routing. Complexity of a protocol is measured by lowering the number of messages to conserve the capacity of the network, from that point of view AODV assures no extra traffic for communications along the existing links.

AODV is invented from the Bellmann-Ford distant vector algorithm. AODV finds a route from a source to a destination only when the source node wants to send one or more packets (traffic) to that destination through several intermediate nodes. The established routes are maintained as long as they are required by the source. It employs the destination sequence numbers to identify the most recent path. A Route Request (RREQ) is flooded throughout the network and it contains the source address or identifier (SrcID), the source sequence number (SrcSeqNum), the destination address or identifier (DestID), the destination sequence number (DestSeqNum), the broadcast identifier (BcastID), and the time to live (TTL) field.

The Broadcast identifier (BcastID) is incremented each time whence the source node sends a new RREQ, so the pair (BcastID, SrcID) identifies a RREQ uniquely. If for the same destination RREQs are received multiple times by a node, the duplicate requests are discarded. When a RREQ is received by an intermediate node, if the intermediate node has either no route for the destination or no up-to-date route, the RREQ will be rebroadcasted with incremented hop count. If a node has a route with a destination sequence number greater than or equal to that of RREQ, a Route Reply (RREP) message is generated and sent back to the source. Every RREQ carries a time to live (TTL) value which indicates that the number of times this message will be rebroadcasted. Every intermediate node, at the time of forwarding a RREQ, enters the previous node address and its BcastID. All intermediate nodes including the destination node having valid routes to the destination are allowed to send Route Reply (RREP) packets to the source. After receiving a Route Reply (RREP) packet, the source node or intermediate node

forwards the data packet to the next node toward the destination. If the source node later receives a RREP consisting of either greater sequence number or the same sequence number with a smaller hop count, it will update its routing information for that destination and starts using the better one. Data packets are buffered locally and transmitted in a FIFO queue when a route is set up. While a node in an active route gets lost, a Route Error (RERR) message is generated to notify the other nodes on both sides of the link about the loss of this link.

Since AODV is a reactive protocol, it uses periodic HELLO messages to inform the neighbors that the link is still alive. The destination sequence number for each destination host is stored in the routing table of a node and it is updated in the routing table when the host receives the message with a greater sequence number. Generally a RREQ is initiated with a small TTL value; gradually it is increased to a certain threshold value for making more efficient route discovery process. In this paper it is proposed to determine the optimum route discovery and the threshold value of TTL by an advanced artificial intelligence based fuzzy neural network in AODV protocol.

There are different types of AODV protocols like Multirate AODV, Reliable AODV, Modified AODV by Swarm Intelligence, AODV-UCSB (University of California, Santa Barbara), AODV-UV (Uppsala University), Kernel-AODV, etc. The connection set up time is lower; at the same time it does not require any central administrative system to control the routing process. AODV reacts fast to the topological changes in the network as it happens for MANET and VANET, and updates only the nodes are affected by these changes. The HELLO messages supporting the routes maintenance are range limited and easy to identify the faults appearing in the routes, so they do not cause unnecessary overhead in the network. AODV also saves storage place, i.e., memory as well as energy or power and bandwidth. It is the best suited for a limited area ad hoc network, since the hop count remains confined within the certain range.

There are some disadvantages also in AODV protocol that determining a reasonable expiry time is difficult one, because the nodes are mobile in MANET and VANET. A route discovery may flood which causes significant network overhead. In larger networks, the nodes may be misbehaved like becoming malicious nodes by attacking the network or uncooperative (selfish) nodes. In AODV it is assumed that all nodes are cooperative such that they help to create route and flow data through the established route.

Furthermore TTL field value is optimized in accordance with the cell structure and average number of nodes lying in the cell. The main advantage of AODV protocol is to create routing path on demand. The best route or path is discovered from the source station to the destination station according to compare different available routes considering their important attributes by applying fuzzy neural network algorithm.

Proposed Artificial Intelligence Based Modified AODV Routing Protocol in VANET

In AODV protocol, a source station (vehicle) initiates a Route Request (RREQ) in the network for connecting to a destination station (node), the route is determined considering the four attributes or parameters like the distance (D), the overload or overhead (O), the consumption of electric power (P), and the expected time (T) to remain the route in alive (active) condition. First three attributes (D, O, P) are acceptable for lesser or minimum value and the fourth attribute expected time is accomplished for larger value, i.e., longer period. Therefore, the normalized expected time period is deducted from one (1) to bring homogeneity among all attributes. There may be different routes under AODV protocol available from the source station to the destination station having one set of values D, O, P, and T for each route. Now the best AODV route is selected among the different routes by applying fuzzy neural network algorithm [16]-[18].

Modified AODV Route Searching by Fuzzy Neural Network

ASA1: The normalized value of a distance (d_1) is equal to D_1/D_{max} , where D_1 is the distance for a particular route, and D_{max} is the maximum available distance among the all routes in the route discovery process. If the normalized route distance for a specific path is taken as d_1 , the membership functions of a fuzzy set F_1 is defined as follows, $\mu_{F_1}(a_1) = d_1$, hence, $F_1 = \{(a_1, d_1)\}$.

ASA2: The normalized value of an Overhead (h_1) is equal to O_1/O_{max} , where O_1 is the overhead or Overload for the specific route, and O_{max} is the maximum available Overhead among the all routes in the route discovery process. If the normalized overhead value is taken as h_1 for the particular route, The membership functions of a fuzzy set F_2 is defined as follows, $\mu_{F_2}(a_2) = h_1$, hence, $F_2 = \{(a_2, h_1)\}$.

ASA3: The normalized value of consumed electric power (p_1) is equal to P_1/P_{max} , where P_1 is the electric power consumed for the route, and P_{max} is the maximum electric power consumed for a route in the route discovery process. If the normalized electric power consumed value is taken as p_1 for the specific route, the membership functions of a fuzzy set F_3 is defined as follows, $\mu_{F_3}(a_3) = p_1$, hence, $F_3 = \{(a_3, p_1)\}$.

ASA4: The normalized value of a time (t_n) available is equal to T_1/T_{max} , where T_1 is the expected time allocated for the specific route, and T_{max} is the maximum available expected time for a route in the route discovery process. Since the normalized available time is preferred for larger value; the normalized unavailable time (t_1) which is favored for lesser value, is taken as the fourth attribute to bring the homogeneity with all other attributes, then t_1 is better for lesser value,

So, $t_1 = (1 - t_n)$. If the normalized time is taken as t_1 for the particular route, the membership functions of a fuzzy set F_4 is defined as follows, $\mu F_4(a_4) = t_1$, hence, $F_4 = \{(a_4, t_1)\}$.

ASA5: Now, the fuzzy operations such as fuzzy set intersection (minimum) and union (maximum) taking three fuzzy membership functions at a time out of total four fuzzy membership functions; the four different values of each fuzzy operation such as fuzzy set intersection or union are obtained as mentioned below:

ASA5.1:

$$T_1 = \mu F_1 \cap \mu F_2 \cap \mu F_3(a) = \min\{\mu F_1(a_1), \mu F_2(a_2), \mu F_3(a_3)\},$$

$$T_2 = \mu F_1 \cap \mu F_2 \cap \mu F_4(a) = \min\{\mu F_1(a_1), \mu F_2(a_2), \mu F_4(a_4)\},$$

$$T_3 = \mu F_2 \cap \mu F_3 \cap \mu F_4(a) = \min\{\mu F_2(a_2), \mu F_3(a_3), \mu F_4(a_4)\},$$

$$T_4 = \mu F_1 \cap \mu F_3 \cap \mu F_4(a) = \min\{\mu F_1(a_1), \mu F_3(a_3), \mu F_4(a_4)\}.$$

ASA5.2:

$$V_1 = \mu F_1 \cup \mu F_2 \cup \mu F_3(a) = \max\{\mu F_1(a_1), \mu F_2(a_2), \mu F_3(a_3)\},$$

$$V_2 = \mu F_1 \cup \mu F_2 \cup \mu F_4(a) = \max\{\mu F_1(a_1), \mu F_2(a_2), \mu F_4(a_4)\},$$

$$V_3 = \mu F_2 \cup \mu F_3 \cup \mu F_4(a) = \max\{\mu F_2(a_2), \mu F_3(a_3), \mu F_4(a_4)\},$$

$$V_4 = \mu F_1 \cup \mu F_3 \cup \mu F_4(a) = \max\{\mu F_1(a_1), \mu F_3(a_3), \mu F_4(a_4)\}.$$

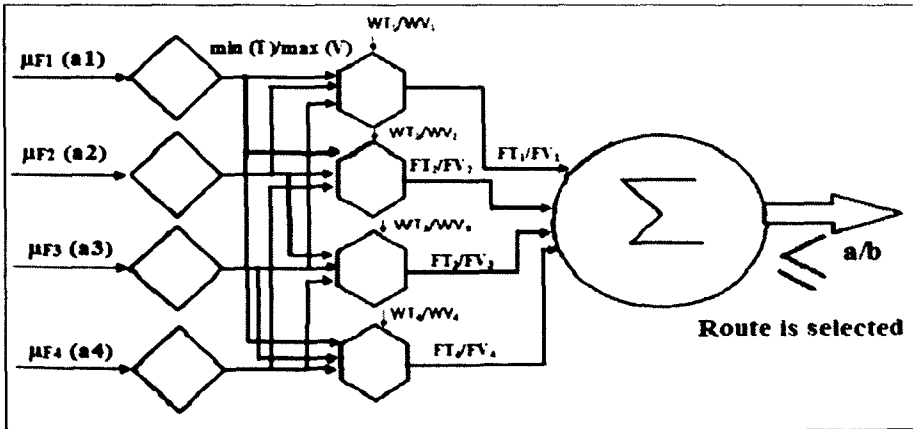


Fig. 6.1: Block diagram of the fuzzy neural network for Modified AODV (M-AODV) Routing Protocol

ASA6: For ascertaining the best or the optimum route under AODV routing protocol, fuzzy neural network algorithm on the results of the fuzzy operations have been applied.

Different weightages to these fuzzy operations (intersection and union) are imposed and these weightages are assigned by altering different values in practical examples and the best values are considered.

$$WT1 : WT2 : WT3 : WT4 = 0.5 : 0.45 : 0.42 : 0.4$$

$$WV1 : WV2 : WV3 : WV4 = 0.9 : 0.85 : 0.83 : 0.8$$

The values of the fuzzy operations are multiplied by the corresponding weightages for computing the optimum or the final values, i.e.,

$$FT1 : FT2 : FT3 : FT4 = T1 \times WT1 : T2 \times WT2 : T3 \times WT3 : T4 \times WT4 \\ = 0.5T1 : 0.45T2 : 0.42T3 : 0.4T4$$

$$FV1 : FV2 : FV3 : FV4 = V1 \times WV1 : V2 \times WV2 : V3 \times WV3 : V4 \times WV4 \\ = 0.9V1 : 0.85V2 : 0.83V3 : 0.8V4$$

ASA7: All the final values of a particular fuzzy operation are defuzzified by a defuzzifying function. Defuzzification is done by the Composite Maxima method, i.e., $\max (FT1, FT2, FT3, FT4) = a$, and $\max (FV1, FV2, FV3, FV4) = b$.

ASA8: The fuzzy-neural rule on the results of the final defuzzified outputs are determined according to examine different values on the practical examples, and then the best suited values are taken.

Thus as per fuzzy neural rule, if $a \geq 0.21$, and $b \geq 0.54$ both satisfies, then it ensures that the route is the best route under M-AODV protocol; otherwise, not. Then the data from the other available routes are to be tested accordingly. Here the value of TTL can be optimized by getting the best route immediately. The block diagram of the fuzzy neural network in selecting the best route under Modified AODV protocol is described in Fig. 6.1.

Implementation Results and Discussion

Example 1: A route under M-AODV protocol is discovered from the source to the destination node with the normalized values of the attributes like distance (D), overhead (O), electric power consumption (P), available expected time (T) as 0.35, 0.44, 0.39, 0.49, respectively; find the suitability of the route.

In this problem, $d1 = 0.35$, $h1 = 0.44$, $p1 = 0.39$, $tn = 0.49$.

Therefore, $t1 = 1 - tn = 1 - 0.49 = 0.51$,

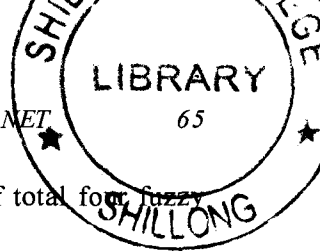
Therefore, $d1 = 0.35$, $\mu F1(a1) = d1 = 0.35$, hence, $F1 = \{(a1, 0.35)\}$.

Therefore, $h1 = 0.44$, $\mu F2(a2) = h1 = 0.44$, hence, $F2 = \{(a2, 0.44)\}$.

Therefore, $p1 = 0.39$, $\mu F3(a3) = p1 = 0.39$, hence, $F3 = \{(a3, 0.39)\}$.

Therefore, $t1 = 0.51$, $\mu F4(a4) = t1 = 0.51$, hence, $F4 = \{(a4, 0.51)\}$.

Now the fuzzy operations such as fuzzy set intersection (minimum) are



computed taking three fuzzy membership functions out of total four fuzzy membership functions.

$$T1 = \mu F1 \cap \mu F2 \cap \mu F3(a) = \min\{\mu F1(a1), \mu F2(a2), \mu F3(a3)\} = \min\{0.35, 0.44, 0.39\} = 0.35$$

$$T2 = \mu F1 \cap \mu F2 \cap \mu F4(a) = \min\{\mu F1(a1), \mu F2(a2), \mu F4(a4)\} = \min\{0.35, 0.44, 0.51\} = 0.35$$

$$T3 = \mu F2 \cap \mu F3 \cap \mu F4(a) = \min\{\mu F2(a2), \mu F3(a3), \mu F4(a4)\} = \min\{0.44, 0.39, 0.51\} = 0.39$$

$$T4 = \mu F1 \cap \mu F3 \cap \mu F4(a) = \min\{\mu F1(a1), \mu F3(a3), \mu F4(a4)\} = \min\{0.35, 0.39, 0.51\} = 0.35$$

Thereafter, we are applying fuzzy neural network algorithm to these fuzzy operations and accordingly the weightages of these fuzzy intersection operations are taken as $WT1 : WT2 : WT3 : WT4 = 0.5 : 0.45 : 0.42 : 0.4$

Now the optimum or the final value is obtained multiplying the fuzzy intersection operation by the corresponding weightage, i.e., $FT1 = T1 \times WT1 = 0.5T1 = 0.5 \times 0.35 = 0.175$

$$FT2 = T2 \times WT2 = 0.45T2 = 0.45 \times 0.35 = 0.1575$$

$$FT3 = T3 \times WT3 = 0.42T3 = 0.42 \times 0.39 = 0.1638$$

$$FT4 = T4 \times WT4 = 0.4T4 = 0.4 \times 0.35 = 0.14$$

Then, all the final values regarding fuzzy intersection operations are defuzzified by the Composite Maxima method in fuzzy neural network, i.e., $\max(FT1, FT2, FT3, FT4) = \max(0.175, 0.1575, 0.1638, 0.14) = 0.175$

Now the fuzzy operations like fuzzy set union (maximum) are calculated taking three fuzzy membership functions at a time out of four fuzzy membership functions.

$$V1 = \mu F1 \cup \mu F2 \cup \mu F3(a) = \max\{\mu F1(a1), \mu F2(a2), \mu F3(a3)\} = \max\{0.35, 0.44, 0.39\} = 0.44$$

$$V2 = \mu F1 \cup \mu F2 \cup \mu F4(a) = \max\{\mu F1(a1), \mu F2(a2), \mu F4(a4)\} = \max\{0.35, 0.44, 0.51\} = 0.51$$

$$V3 = \mu F2 \cup \mu F3 \cup \mu F4(a) = \max\{\mu F2(a2), \mu F3(a3), \mu F4(a4)\} = \max\{0.44, 0.39, 0.51\} = 0.51$$

$$V4 = \mu F1 \cup \mu F3 \cup \mu F4(a) = \max\{\mu F1(a1), \mu F3(a3), \mu F4(a4)\} = \max\{0.35, 0.39, 0.51\} = 0.51$$

Weightages of this fuzzy union operations are as $WV1 : WV2 : WV3 : WV4 = 0.9 : 0.85 : 0.83 : 0.8$

The optimum or the final values regarding fuzzy union operation applying

fuzzy neural network algorithm are below: $FV1 = V1 \times WV1 = 0.9V1 = 0.9 \times 0.44 = 0.396$

$$FV2 = V2 \times WV2 = 0.85V2 = 0.85 \times 0.51 = 0.4335$$

$$FV3 = V3 \times WV3 = 0.83V3 = 0.83 \times 0.51 = 0.4233$$

$$FV4 = V4 \times WV4 = 0.8V4 = 0.8 \times 0.51 = 0.408$$

All the final values are defuzzified by the Composite Maxima method which yields, i.e., $\max(FV1, FV2, FV3, FV4) = \max(0.396, 0.4335, 0.4233, 0.408) = 0.4335$,

Now applying fuzzy-neural rule, $\max(FT1, FT2, FT3, FT4) = 0.175$, i.e., $? 0.21$ and $\max(FV1, FV2, FV3, FV4) = 0.4335$, i.e., $? 0.54$, therefore, the route detected under M-AODV scheme is accepted and may be used for traffic (data) flow. Thus, TTL must possess the least value, as the best route is obtained.

A. Simulation

In this study, we used NCTUns-6.0[12] for simulation. We have chosen this simulator because,

- A. Highly integrated and professional GUI environment.
- B. Support for various network protocols.
- C. Support for various important network.
- D. Same configuration and operations as for real life networks.
- E. High simulation speed and repeatable simulation result.
- F. High fidelity simulation results.

Performance Metrics

Different performance metrics are used to check the performance of routing protocols in various network environments. In our study we have selected throughput and packet drop to check the performance of VANET routing protocols against each other. The reason for the selection of these performance metrics is to check the performance of routing protocols in highly mobile environment of VANET. Moreover, these performance metrics are used to check the effectiveness of VANET routing protocols i.e. how well the protocol deliver packets and how well the algorithm for a routing protocol performs in order to discover the route towards destination. The selected metrics for routing protocols evaluation are as follows [8,9].

1. Throughput

Throughput is the average number of successfully delivered data packets on a communication network or network node. In other words throughput describes as the total number of received packets at the destination out of total transmitted

packets [8]. Throughput is calculated in bytes/sec or data packets per second. The simulation result for throughput in NCTUns6.0 shows the total received packets at destination in KB/Sec, mathematically throughput is shown as follows:

Throughput (bytes/sec) =

Total number of received packets at destination* packet size

Total simulation time

2. Packet Drop

Packet drop shows total number of data packets that could not reach destination successfully. The reason for packet drop may arise due to congestion, faulty hardware and queue overflow etc. Packet drop affects the network performance by consuming time and more bandwidth to resend a packet. Lower packet drop rate shows higher protocol performance.

3. Collision

The Collision of data packet is the number of packets collides to each other due to congestion. It affects the performance directly on the bandwidth. Lower packet collision rate shows higher protocol performance.

Table 6.1: Data Set Generation

D	O	P	T
0.32	0.41	0.54	0.40
0.32	0.45	0.58	0.43
0.331	0.45	0.50	0.54
0.32	0.52	0.42	0.56

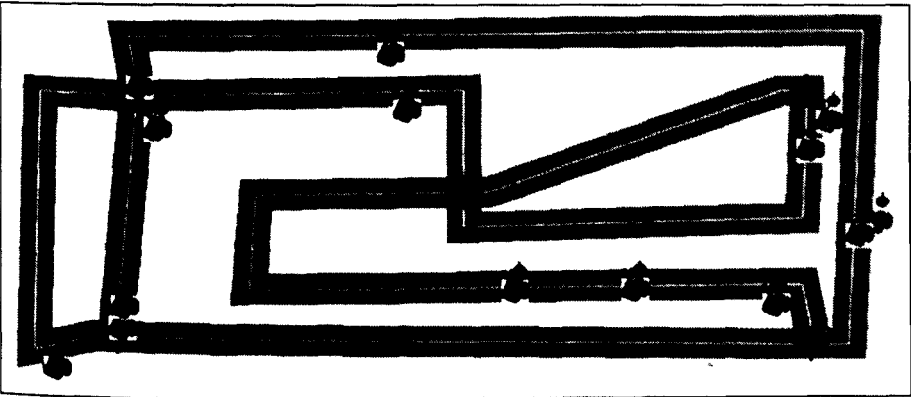


Fig. 6.2: Testing Scenario

B. Testing Parameters**Table 6.2: AODV Testing Parameters**

Parameter	Settings
Transmission mode	TCP/IP
Lane Width	20m
Simulation time	400sec
RTS threshold	3000bytes (O)
The car profile (Taken five)	18km/H, 36km/H, 50km/H, 60km/H, 80km/H
Number of lane	2
The protocol	AODV
standard used for each vehicular node	IEEE802.11b
cars are selected for three different scenarios	10,15,20,25,30 (D)
Transmission power used	15dbm (P)
TTL	7(T)

Table 6.3: aiAODV Testing Parameters

Parameter	Settings
Transmission mode	TCP/IP
Lane Width	20m
Simulation time	400sec
RTS threshold	3000bytes (O)
The car profile (Taken five)	18km/H, 36km/H, 50km/H, 60km/H, 80km/H
Number of lane	2
The protocol	AODV
standard used for each vehicular node	IEEE802.11b
cars are selected for three different scenarios	10,15,20,25,30 (D)
Transmission power used	15dbm (P)
TTL	7(T)

C. Results

CAR 10:

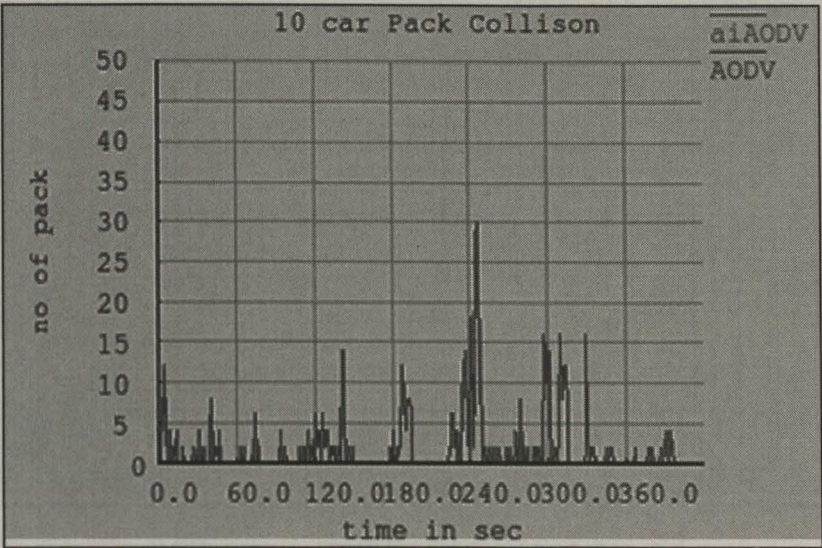


Fig. 6.3: Number of Packet in collision in AODV vs aiAODV

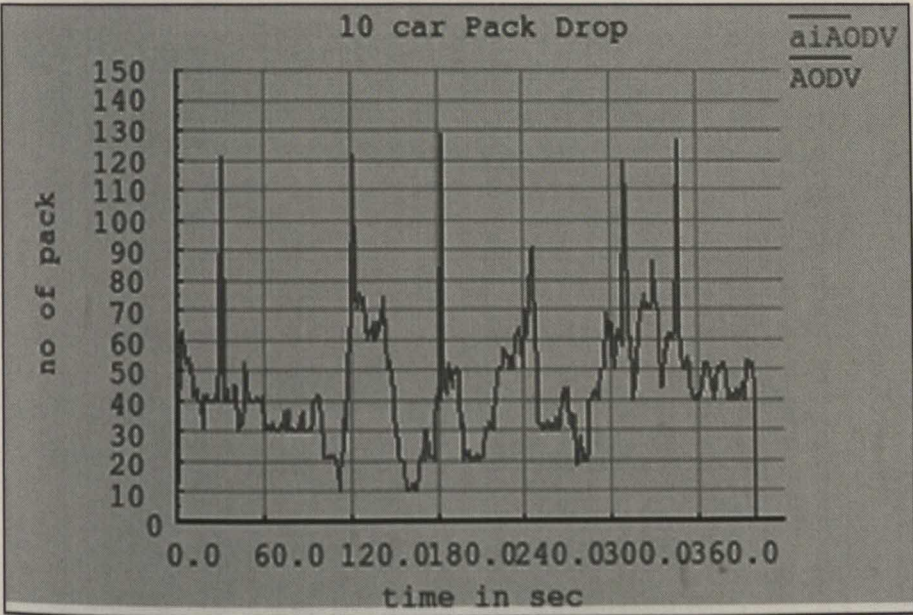


Fig. 6.4: Number of Packet drop in AODV vs aiMAODV

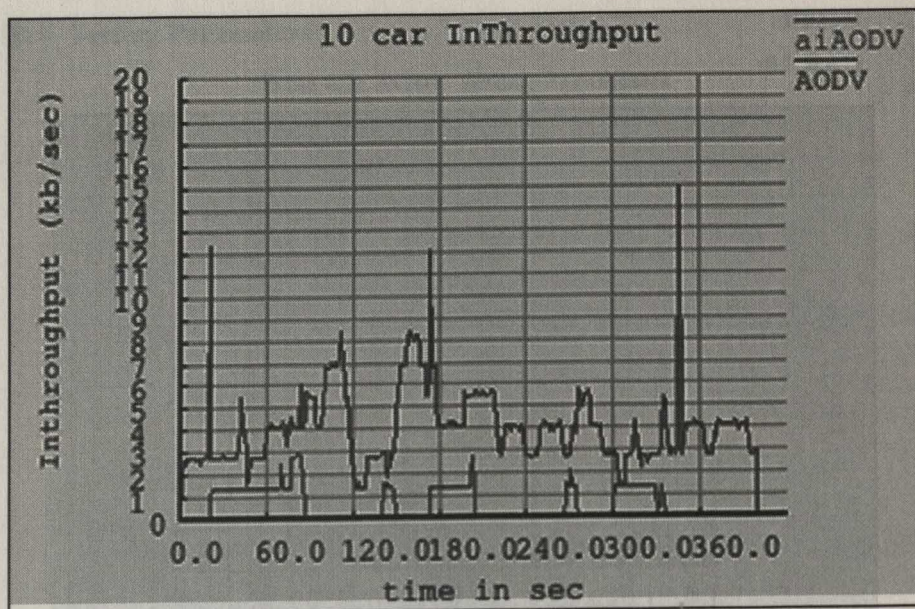


Fig.6.5: In throughput of AODV vs aiMAODV

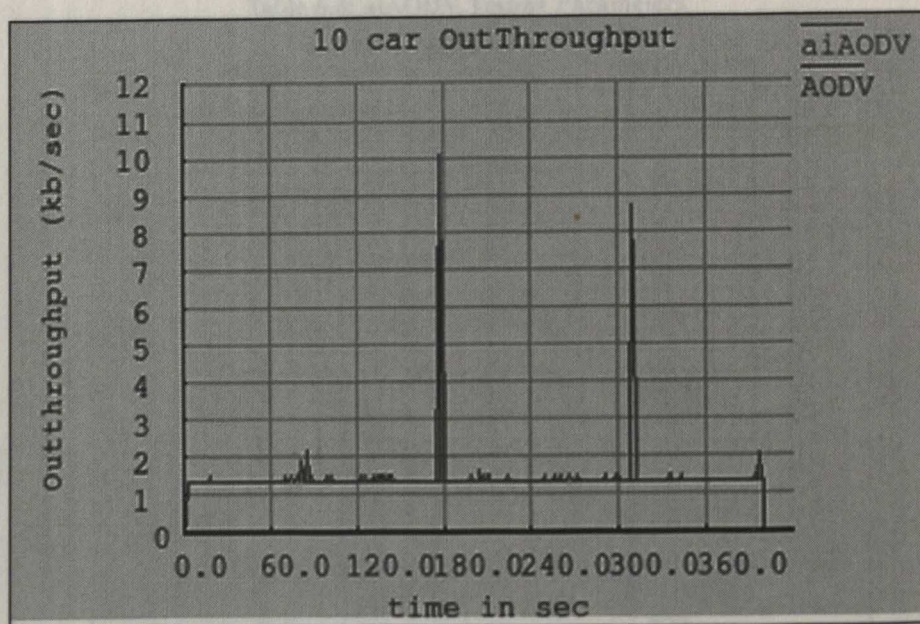


Fig. 6.6: Out throughput of AODV vs aiMAODV

CAR 15:

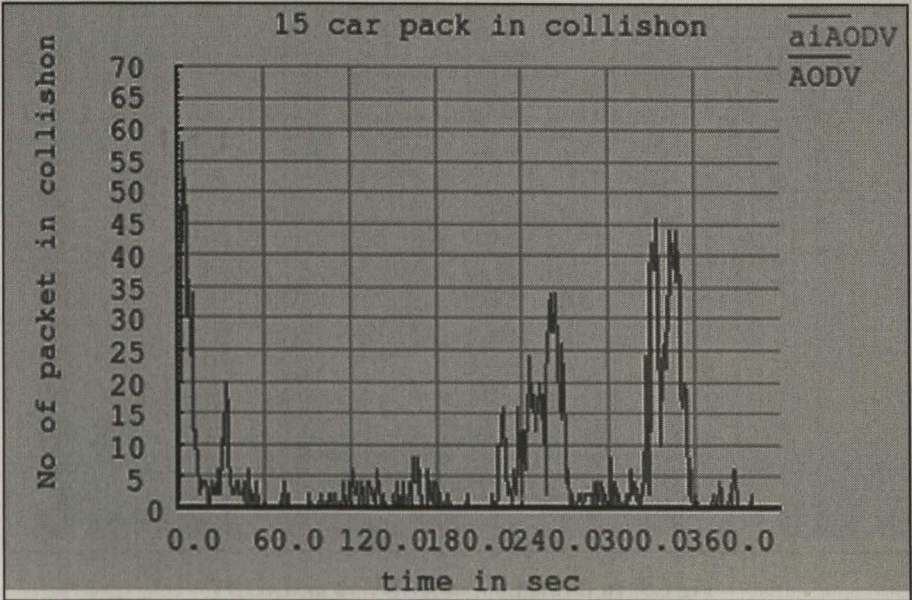


Fig.6.7: Number of Packet in collision in AODV vs aiAODV

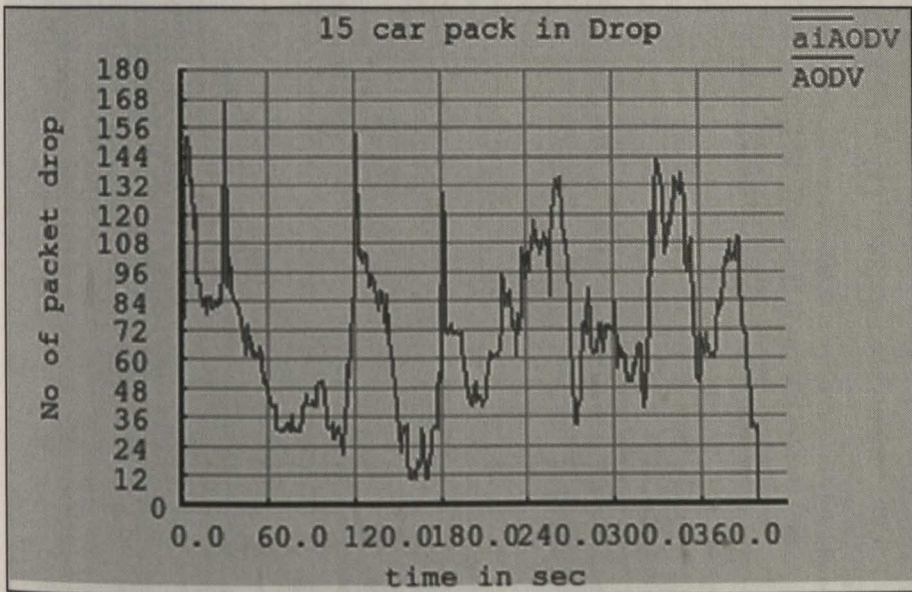


Fig. 6.8: Number of Packet drop in AODV vs aiMAODV

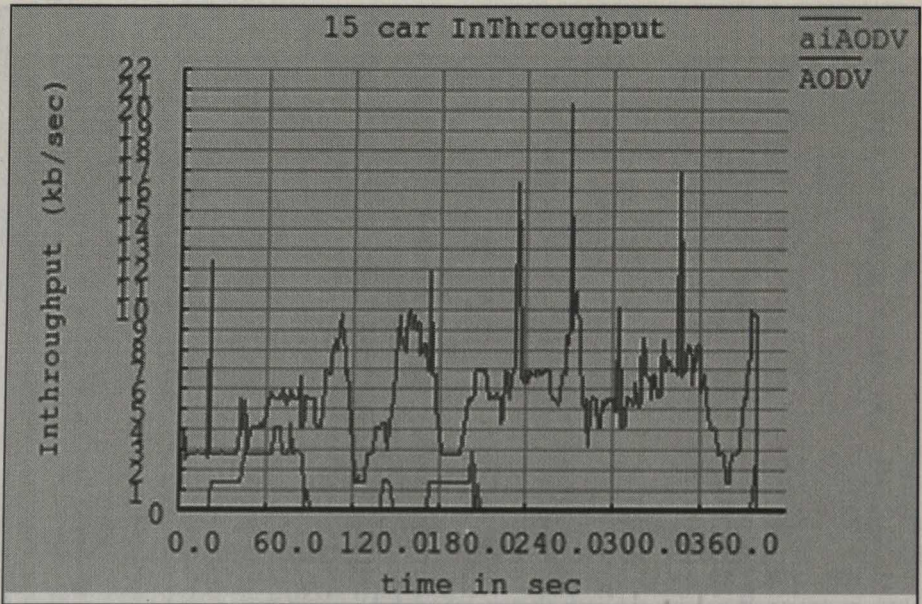


Fig.6.9: In throughput of AODV vs aiMAODV

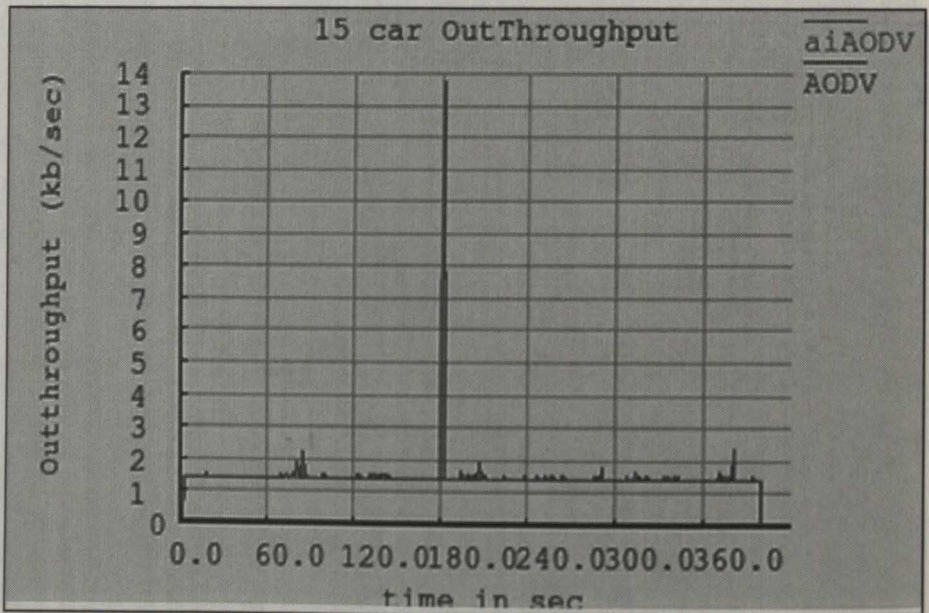


Fig. 6.10: Out throughput of AODV vs aiMAODV

CAR 20:

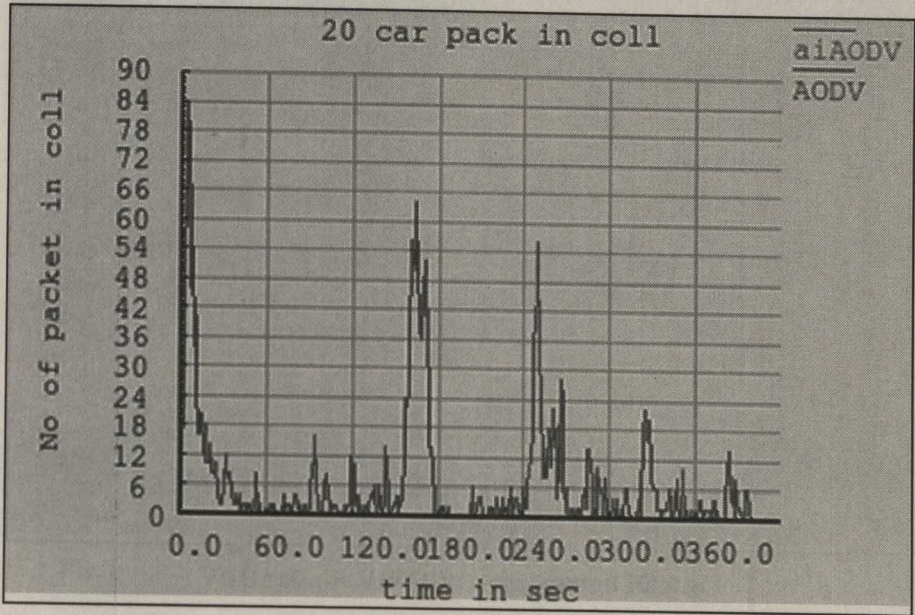


Fig.6.11: Number of Packet in collision in AODV vs aiAODV

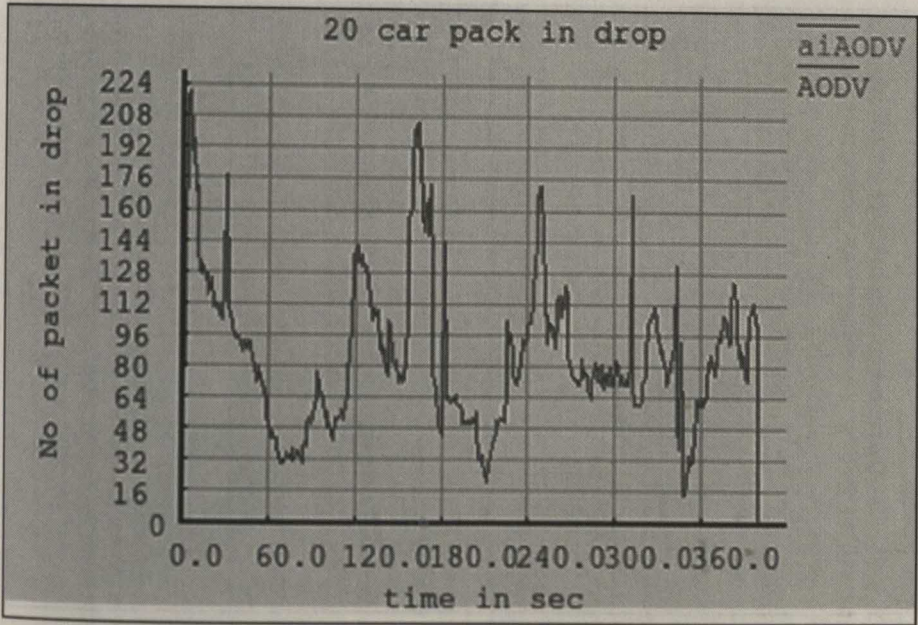


Fig. 6.12: Number of Packet drop in AODV vs aiMAODV

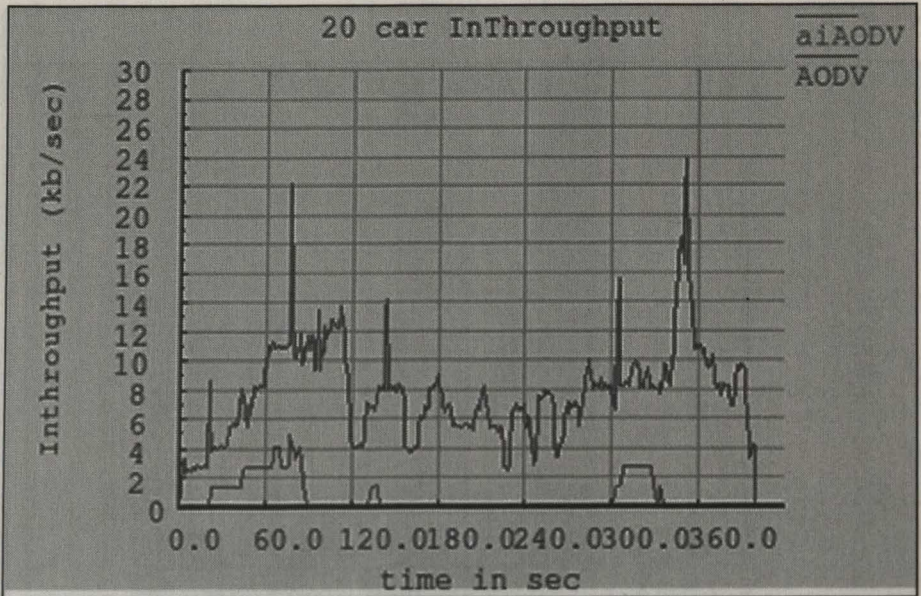


Fig.6.13: In throughput of AODV vs aiMAODV

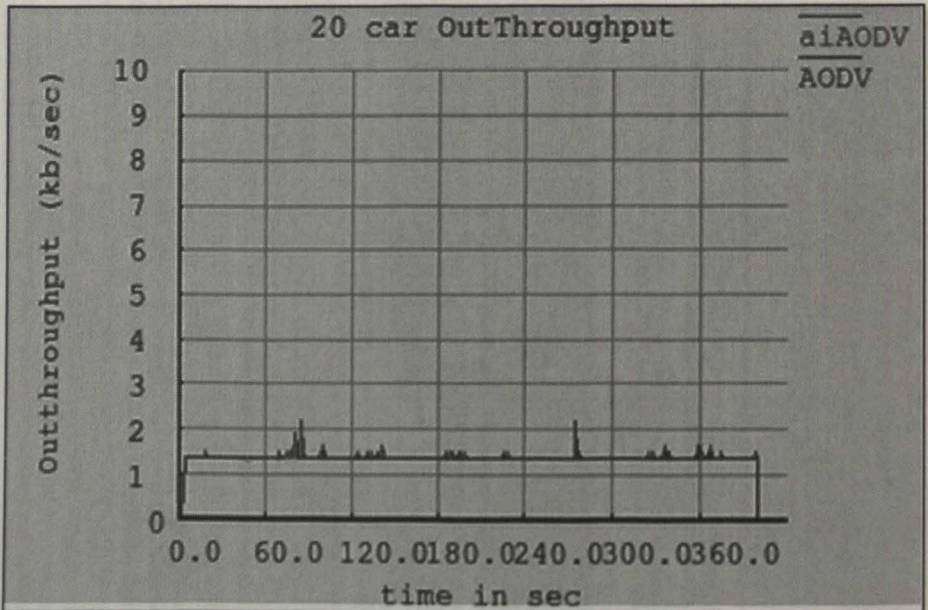


Fig. 6.14: Out throughput of AODV vs aiMAODV

CAR 25:

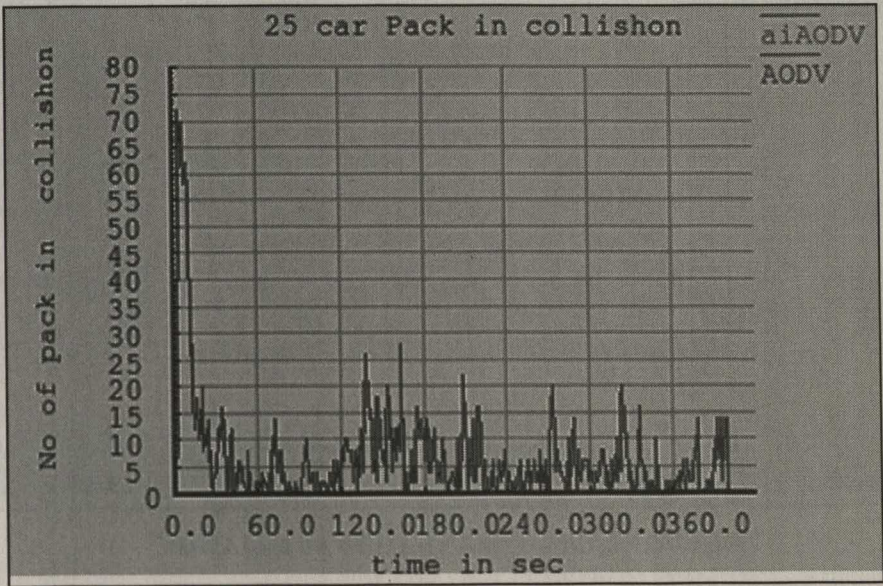


Fig.6.15: Number of Packet in collision in AODV vs aiAODV

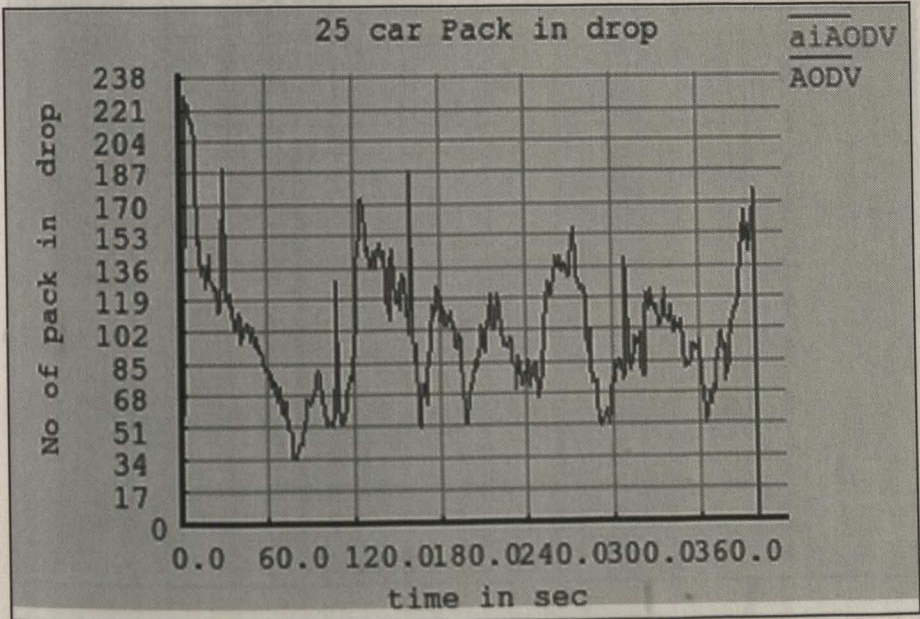


Fig. 6.16: Number of Packet drop in AODV vs aiMAODV

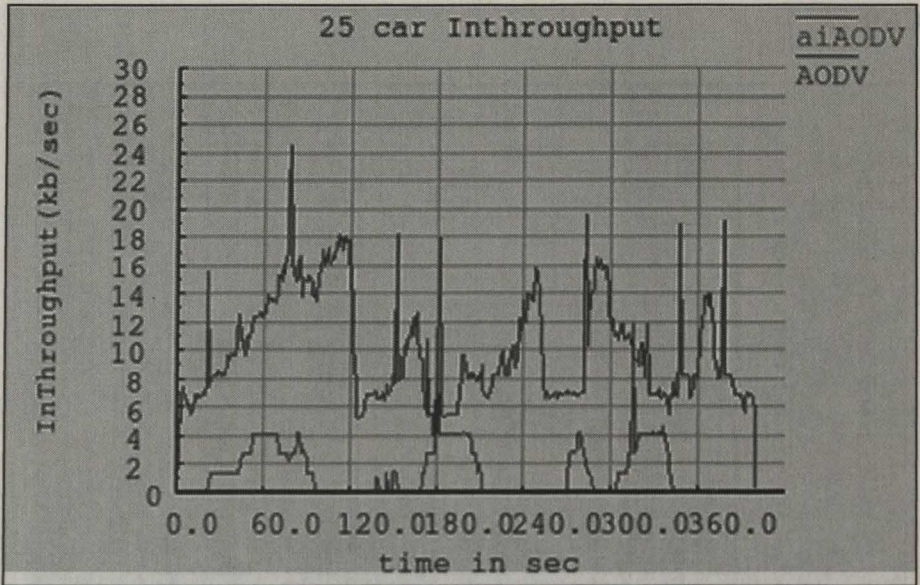


Fig.6.17: In throughput of AODV vs aiMAODV

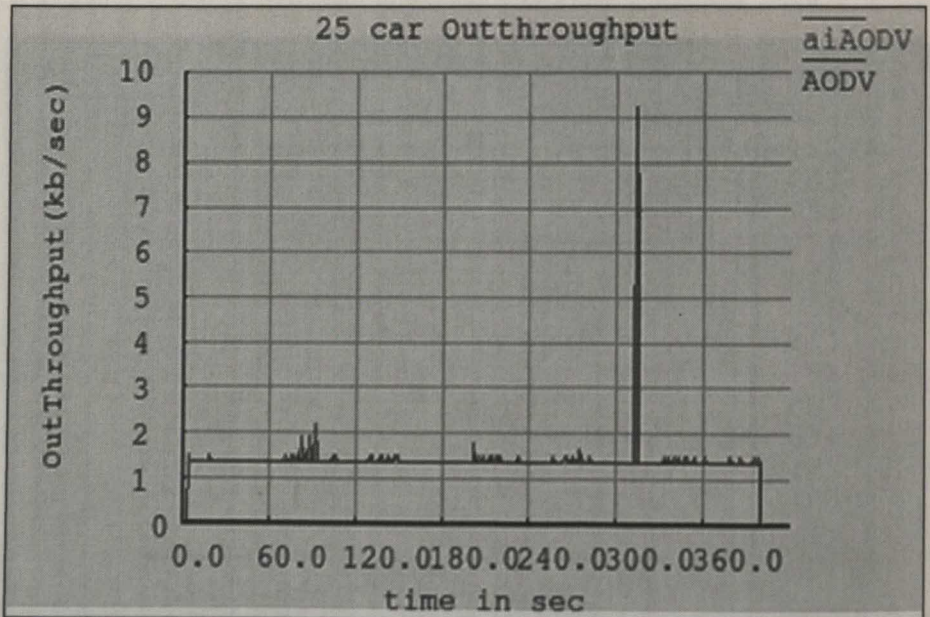


Fig. 6.18: Out throughput of AODV vs aiMAODV

CAR 30:

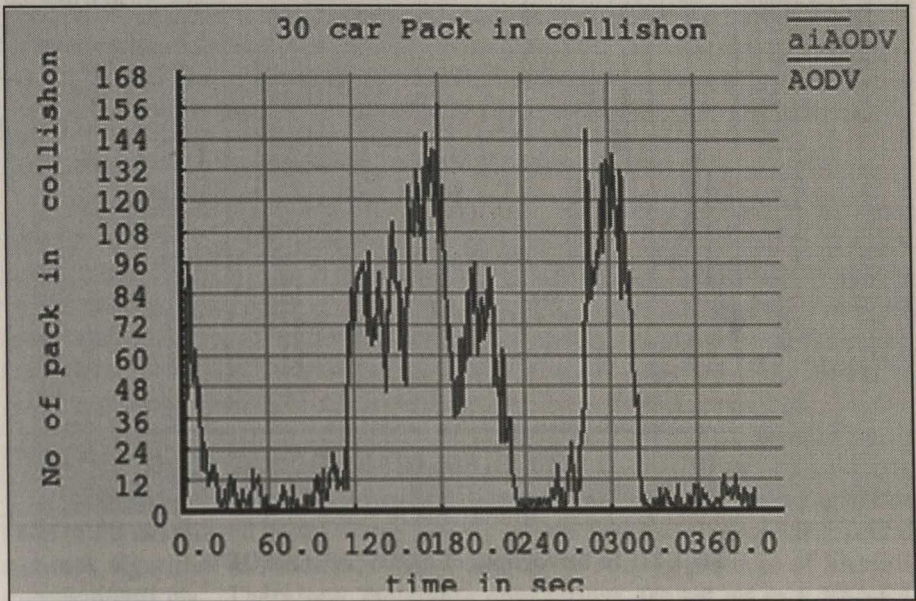


Fig.6.19: Number of Packet in collision in AODV vs aiAODV

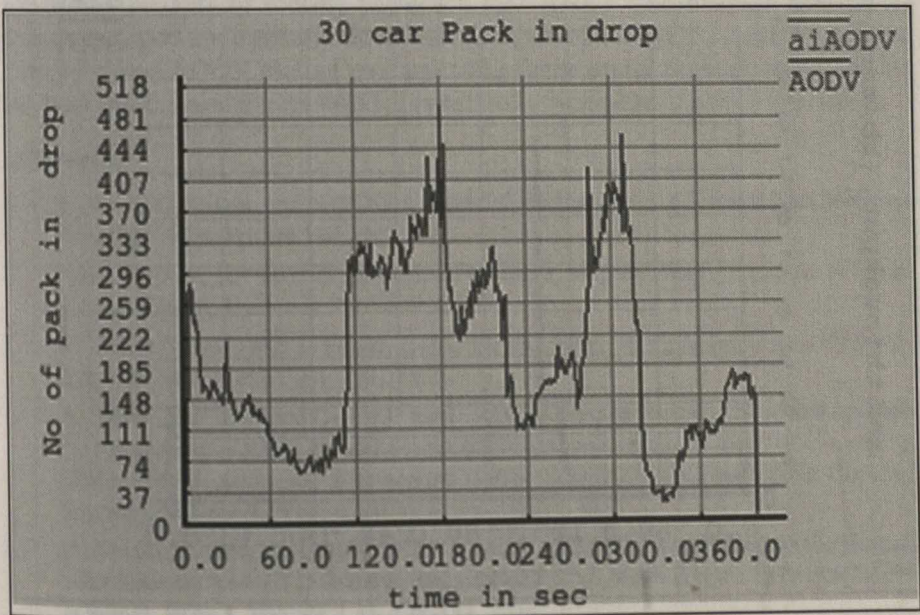


Fig. 6.20: Number of Packet drop in AODV vs aiMAODV

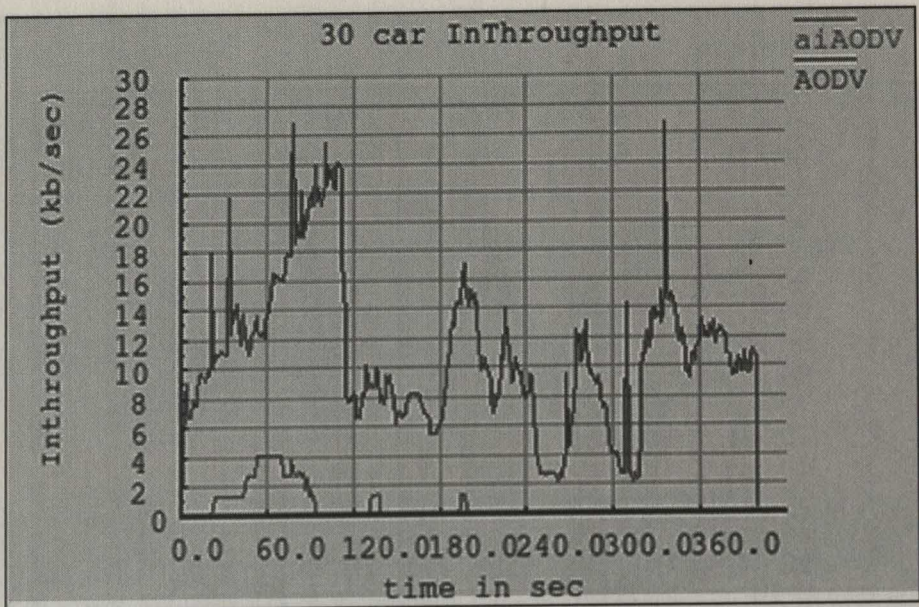


Fig.6.21: In throughput of AODV vs aiMAODV

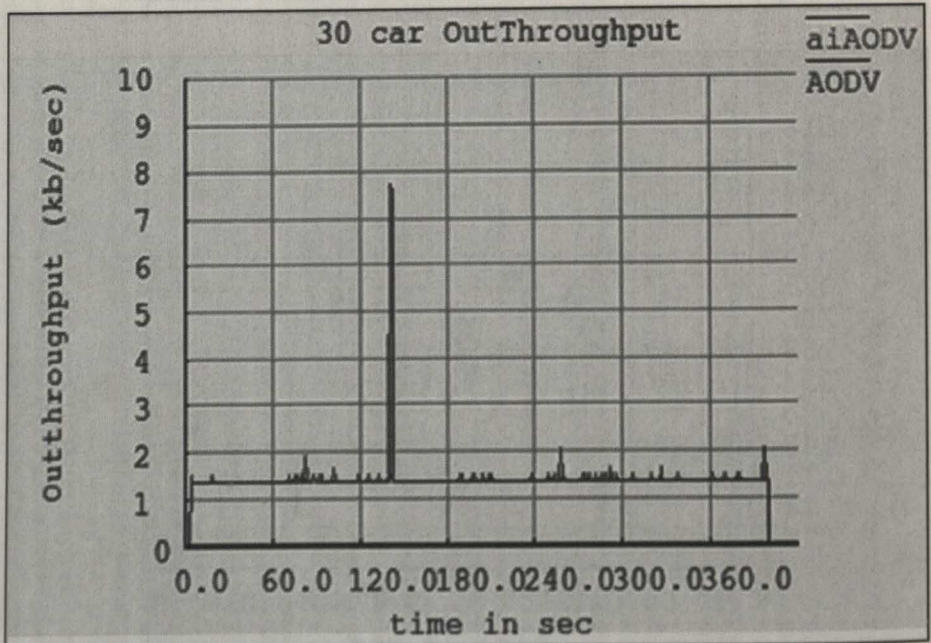


Fig. 6.22: Out throughput of AODV vs aiMAO

We found the packet drop and collision is drastically reduced. In-throughput (packet incoming to node) is also diorite from original AODV as number packet to discover is very less in our proposal. But, the out-throughput is remaining same. It indicates with reduced Power (P), Overhead (D), TTL (T) and same distance (D) we got the satisfactory result. That proves the optimal parameters usage to discover the routing path is successful according to our proposal using neural network

Advantages of the Modified AODV Routing Protocol

This route discovery technique under Modified AODV protocol is the most efficient one due to applying artificial intelligence [AI] in advanced stage, i.e., fuzzy neural network used. Also it does not require any further information to supply by the source (host) node while making a call. So it is a unique one. The route is determined by the essential attributes or parameters of the AODV protocol. No cryptography algorithm or any complex functions are applied. This Modified AODV protocol technique ensures the fastest best route discovery.

Conclusions

In this paper we have proposed to determine the best route from the source node to the destination node through several intermediate nodes using fuzzy neural network algorithm in Modified AODV (aiAODV) routing protocol of VANET communications. Therefore, the stable connections are set up in a VANET communications by implementing a fast and easy routing techniques like artificial intelligence based Modified AODV routing protocol in the VANET system. There are many variants of routing protocols for VANET transmission have been proposed, those are basically the modified forms of MANET routing protocols. This Modified AODV routing protocol using fuzzy neural network algorithm for the best route searching in VANET is fantastic workable in a real time basis.

References

1. E. Rich, K. Knight, and S. B. Nair, Artificial Intelligence, 3rd Edition, Tata McGraw Hill Education Private Ltd., 2010.
2. Vilem Novak, Jiri Mockor, and Irina Perfilieva, Mathematical Principles of Fuzzy Logic, Kluwer Academic Publisher, 2006.
3. J. Yan, M. Ryan, and J. Power, Using Fuzzy Logic: Towards Intelligent Systems, Prentice-Hall of India Pvt. Ltd., 1995.
4. S. Saha, Dr. U. Roy and Dr. D.D. Sinah, AODV Routing Protocol Modification with Broadcasting RREQ Packet in VANET: International Journal of Emerging Technology & Advanced Engineering (ISSN 2250-2459, ISO 9001:2008 Certified Journal), Volume 4, Issue 8, pp 439-444, August 2014.
5. A. Modak, S. Saha, Dr. U. Roy, Dr. D.D. Sinah, AODV Routing Protocol Modification with Stack for VANET INin City Scenarios: IJETCAS, ISSN (Print): 2279-0047 ISSN (Online), Issue 8, Volume 2, pp 128-133, 2014.
6. RFC of AODV, DSR: www.ietf.org/rfc/rfc3561.txt, www.ietf.org/rfc/rfc4728.txt.

7. S.Y. Wang, C.C. Lin, and C.C. Huang: Nctuns tool for evaluating the performances of real-life p2p
8. Applications, In Nova Science Publishers, Inc., ISBN 978-1-60876-287-3, pp. 1-23, 2010.
9. P. K. Bhattacharjee and U. Roy, "AODV Routing Protocol Is the Most Useful Routing Protocol in VANET Communications" National Conference on Computing and Systems-2012 (NaCCS-2012), Computer Science Department, Burdwan University, Mar, 2012, pp. 1-6.
10. P. K. Bhattacharjee and U. Roy, "Modified AODV Routing Protocol Explored by Swarm Intelligence Technique in VANET Communications" IEEE International Conference in Mobile and Embedded Technology (MECON-2013), Amity University, Delhi, India, pp. 235-242, January 2013.
11. S. Saha, Dr. U. Roy, Dr. D.D. Sinah, Sk. A. Ahmed: Performance Analysis of VANET Scenario in Ad-hoc Network by NCTUns Simulator. published in INTERNATIONAL CONGRESS On "Innovative Trends in Information Technologies and Computing Sciences for Competitive World Order"(ITITCSCWO - 2013), New Delhi , JNU, pp 2-3, March 2013.
12. S. Saha , Dr.U. Roy, Dr. D.D. Sinah : Performance comparison of various Ad-Hoc routing protocols of VANET in Indian city scenario, published in AIJRSTEM 14-126 ,ISSN (Online): 2328-3580,pp 49-54, March 7, 2014.
13. NCTUns6.0 protocol developer manual; <http://elearning.vtu.ac.in/15/E-Notes/NW%20prog%20lab/NCTUns%20Manual.pdf>

Recent Development and Challenges In Mobile Ad Hoc Network

*Indrani Das
Bipul Shyam Purkayastha
Sanjoy Das*

Abstract

Recent rapid developments of wireless domain attract researchers, academicians and others to explore them in more details. A Mobile Ad hoc Network is a collection of mobile devices equipped with wireless technology and is capable to communicate with each other through wireless medium. This type of network does not require any fixed infrastructure for communication. This striking feature helps its quick and easy deployment in a situation where fixed infrastructure may not be possible to be set up or established. Although, this network has these attractive features but still there are lot of challenges and issues left and there are needs to explore for better understanding of the network. In this paper, we have presented various challenges and current issues existing in Mobile Ad hoc Networks. We expect new researchers will be benefited through this paper.

Keywords: *MANET, Throughput, dynamic, topology, power.*

Introduction

A Mobile Ad-hoc Network (MANET) is a cooperative engagement of collection of mobile hosts without intervention of a centralized infrastructure. A MANET is a kind of mobile wireless network in which every node is mobile and autonomous [1, 2]. These mobile devices are capable to communicate with each other through wireless media. As this type of network does not require any fixed

infrastructure, so deployment cost incurred is very low and time efficient. The wide application areas of an ad hoc network are battlefield communications, emergency rescue missions, law enforcement, riot control, spying, deep rural areas and convention centres etc. This network may provide a cost-effective way of sharing information among many mobile hosts. These are the unique characteristics of MANET that differentiate it from other types of wired and wireless networks.

In this network, the wireless devices are of limited transmission range; therefore, the routes between a given pair of nodes often consist of multiple hops. The nodes in the network cooperate to relay data packets towards destinations. The devices are often powered through batteries and as a result the depletion of battery power may cause failure of nodes as well as links.

In section 2, various challenges related to this network are described. Section 3 shows the future scopes in this network. Finally, the paper is concluded in section 4.

Challenges in Mobile Ad-hoc Networks

The node mobility in this network is unpredictable; they may join and leave the network any time. Network topology is very dynamic in nature due this frequent disconnection which occurs in the network. The possible reasons are that nodes may go away from coverage area of each other; a node may die due to complete depletion of battery power. Following are the challenging issues [1,3] which need to be addressed by new researchers in designing any protocol for MANET so that this type of network may get more popularity.

i. Dynamic Topology[4]

Unpredictable node movements with varying node speed increase frequent fragmentation in the network. So, data delivery to a destination becomes a challenging issue.

ii. Battery Power [4]

All mobile devices or nodes are equipped with limited battery power. The whole network life time fully depends on the battery power of individual node. If any node's battery power depletes completely it creates link disconnection in the network and data delivery becomes a challenging task.

iii. Bandwidth utilization[4]

Bandwidth is a very scarce resource in this network. Bandwidth utilization should be done properly and economically.

iv. Quality of service[4]

The evaluation of any network is measured through QoS parameters like

average end-to-end delay, packet delivery ratio, throughput, routing overhead, path length, hop count and many more. So it is always desirable that packet delivery ratio [5], and throughput of the network should be high, and delay, path length, hop count should be minimum. To achieve these things efficient management of all the resources of the network is required.

v. Node cooperation

It is very seldom that source and destination node fall in direct transmission range of each other. Data can only be delivered when two or more nodes exist in the other's direct transmission range. Here, the time during which the two nodes are connected is very less, so node cooperation is very much needed to establish multi hop communication and then data delivery becomes successful.

vi. Security [6]

This is the most challenging issue in this network, because to authenticate whether a node is malicious or not, consumes many network resources. If any malicious node exists in the network, it creates havoc for the network. So robust mechanism to detect and identify such malicious node is always needed.

Future Scope

For the last few decades various developments in different domains of Mobile Ad hoc Networks have occurred. Some of the recent developments or thrust areas of this network where further improvements may be possible are efficient power management, designing good MAC protocol, efficient routing mechanism, more security and better load balancing etc. In the context of various challenges discussed in previous section the following parameters can be further explored:-

- i. **Routing Mechanism:** Efficient routing mechanism is needed for guaranteed and timely delivery of data to the destination.
- ii. **Energy conservation:** Mobile battery power should be used in an optimized way, so that network life time can be improved further.
- iii. **Security:** Node authentication is a very essential component of the network. Otherwise, it may mislead the other nodes in the network to act according to that node. If such node exists in the network undetected for longer time, it will lead to complete disruption of the network. A robust mechanism is needed to protect the network from black hole attack, worm hole attack ,DDoS attack etc.
- iv. **Load balancing:** The network is cooperative in nature. If any node is over loaded with traffic, then that should be distributed to other nodes in an efficient way so that node life time can be improved, and network life time can be increased.

Conclusion

The objective of the paper is to help new researchers to explore more and more areas of Mobile ad hoc network. We have discussed various issues that exist in the network and have also shown various domains where further improvement may be possible. There may exist many more areas which need to be further explored. Finally, for deployment of any ad hoc network, it should be reliable, secure and resource utilization should be in an efficient way. Otherwise, acceptability of such network will be low. Therefore, while designing a routing protocol, or for that matter any other protocol these basic issues must be addressed.

References

1. C. Siva Rama Murthy and B. S. Manoj, "Ad Hoc Wireless Networks Architectures and Protocols", Second Edition, Pearson, 2008.
2. Ram Ramanathan and Jason Redi, "A Brief Overview of Ad Hoc Networks: Challenges and Directions", *IEEE Communications Magazine*, Vol. 40, No. 5, pp. 20-22, May 2002.
3. C K Toh, *Ad Hoc Mobile Wireless Networks, Protocols and Systems*, PHI, 2002.
4. Imrich Chlamtac, Marco Conti, and Jennifer J.-N. Liu, "Mobile ad hoc networking: imperatives and challenges", *Ad Hoc Networks*, Vol. 1, 2003, pp. 13-64.
5. Mukesh Kumar et al., "Issues and Challenges of Quality of Service in Mobile Ad-hoc Network", *International Journal of Computer Science & Engineering Technology (IJCSET)*, Vol.3, No.1, pp.61-66.
6. Hao Yang, HaiyunLuo, Fan Ye and Songwu Lu "Security in mobile ad hoc networks: challenges and solutions", *IEEE Wireless Communications*, Vol.11, Issue: 1, pp.38 - 47, 2004.

Cyber Space and Its Theories

Arjun Chetry
Moutan Sarkar

Abstract

The danger of Cyber Warfare ranks second only to that of Nuclear War. But Cyber Warfare is more ticklish than any other form of war, it is completely silent, partially hidden, highly heavy and fully curtailed. Proliferation and threat of nuclear weapons is comparatively easy to manage. Cyber Warfare is a strategic threat in Information age. The unseen power of the cyber space has already started showing up its ugly face. United States already treats the importance of Cyber Security at the similar levels as economic and military Security. It's high time that not Cyber War but at least Cyber Security be the Nation's Action plan. This paper discussed the issues & security measures.

Keywords: Cyber crime, hacking tools/techniques, information security, online safety, cyber crime investigation.

Introduction

To spend a important part of our day on the Internet for many of us is quite normal. As with any life-changing force so is the case with internet. And considering the management and quality of life, it is extremely difficult to control its misuse. Let me take you through a simple case. Most of the celebrities say Justin Bieber, Adam Sandler, Aretha Franklin, Charlie Sheen, Bill Cosby, Lindsay Lohan, Nick Jonas and Taylor Swift and also Amitabh Bachchan, Sachin Tendulkar, Yuvraj Singh, Honey Singh they all have been declared dead in the Social media at some point of time. And there are umpteen such illustration of

rumors in the social media. And not only celebrities but also Corporate Biggie have been a victim of all this. On a wider spectrum there have been attacks on Countries Critical Information Infrastructure, and the attacks were not always carried on by Script Kiddies but they were funded by National Bodies or other alarming trunk. The capabilities and their consequences of Cyber Warfare is not easily grasped.

Issues and Perspective of Information Security

In internet, cyber criminals have widened their crime range & the percentage of successful investigation against those crimes for tracing cyber criminal is too poor. The technique used to compromise the user computers/applications always varies from person to person but the major steps are information gathering, scanning, gaining control, maintaining access & clearing tracks. There is no need of person being technical due to the availability of complex software with very simple GUI freely over internet.

The use of social engineering techniques to initiate criminal activities becomes untraceable due to availability of platforms like proxies, tor browser, Virtual Private Networks, etc using which they can distribute the malwares & get the control of any remote systems & remain anonymous over internet.

Simple Google dorks/scanning techniques are successful in retrieving information of systems or servers without any expert knowledge. The tools used for this purposes are the same tools e.g. Nmap, Acunetix, Metasploit framework, aircrack-ng, tcpdump, wireshark, etc which are being used by the security researcher/pen tester across the globe.

Similar hacking techniques in case of smart phones are increasing with greater pace then in case of computer systems. In case of Android there are various cases over internet regarding malwares which cheating persons & retrieving their information. If we check the statistics, then the number of malwares in case of hand held devices are much more as compare to computer system. Along with the threats from third party apps, there are many cases reported for the apps that are available in play store.

Social networking site is one of the best places for cyber criminal to catch fish from the ocean called as internet. The range of person using this service is from core technical person to laymen it itself is challenge for the service provider in providing security. The analysis of social networking site can reveal the answers of thousands of questions about the person. Another issue of social networking site is that the person can easily cheat & become the friend of targeted person & start doing the first step, i.e. information gathering of hacking. There are occasions where data leakage from social networking site was reported & leaving a question against the security of server & showing danger of keeping

private data even after having proper privacy setting. Cases initiated through the chat of such sites which later leads to murder, suicide, defamation, etc are reported number of times & many registered cases are also there.

One more issue of online servers whether social networking or cloud or any other sites, there may be a case that their server got compromised leading to loss of entire data of their account holders. Here the questions of data privacy comes & why service provider not taken proper measures to keep the data of customers safe.

Reported issues as reported/warns by various organizations/sites

- 1) Google news given a warning that security researchers FireEye on 10/11/2014 reported a vulnerability of iOS systems version ios 7.1.1 & later that the system can be replace with fake applications & it can be use to install malware or perform other activities in it.
- 2) FBI: We face cyber threats from state-sponsored hackers, hackers for hire, global cyber syndicates, and terrorists. They seek our state secrets, our trade secrets, our technology, and our ideas- things of incredible value to all of us. They seek to strike our critical infrastructure and to harm our economy. Src: www.fbi.gov on September 10 2014.
- 3) According to <http://thehackernews.com>, Researcher Mordechai Guri, along with Professor Yuval Elovici of Ben Gurion University, presented the research in the 9th IEEE International Conference on Malicious and Unwanted Software (MALCON 2014) held at Denver, that the FM radio signals can be used using malware to lift data from a system which is completely isolated from the internet or any wifi connection.

Investigation Issues

Internet is changing day by day with the modification of more and more complex infrastructure from time and again introducing new challenges to law enforcement agencies of the world. Every crime committed using technology or to technology is of elusive nature so there is a need of development or enhancement of the knowledge of law enforcement personnel to make the investigation possible as well as there is a huge requirement for general public awareness so that they can take preventive measures. There are various factor which leads to tracing or investigation failure in this virtual world, Internet.

The procedure of investigation of cyber crime is different from the investigation of traditional crime. For example, in case of traditional crime scene investigating officer (IO) needs to wear gloves to preserve finger print during evidence collection but in case of digital devices of crime scene IO have to also preserve the data inside the device along with finger prints. Hence, educating law

enforcement personnel as well as general public about the nature of this crime is very important.

Along with such issues, there are various difficulties due to advancement of technologies & unavailability of software/tools/techniques or professional. Few of the major problems for cyber crime investigation are listed below:

- i. Availability of open proxies: The criminal can use proxies to hide their locations.
- ii. Availability of encryption techniques: Interception of communication of cyber criminal is almost impossible due to availability of lots of encryption techniques.
- iii. Use of Tor browser: Use of tor browser bundle gives more and more challenges for tracing as the packets can be routed through multiple countries before delivering to concern person.
- iv. Virtual private Network (VPN): The usefulness of VPN is known to every person but it also gives challenges for law enforcement whenever the criminal use VPN for any criminal activities.
- v. Virtual Machine: Use of virtual machine can lead can be one more hurdle for performing investigation.

Responsibility of Users

To prevent cyber crime completely may not be possible but we can safeguard or make it difficult for criminal mind if we take some preventive measures. Some of the common measures are:

1. In case of online activities

- i. Updating or learning about the criminal activities & their recent techniques can be the best method to safeguard them. Hence, cyber awareness program or knowledge should be adopted by every individual.
- ii. Click with caution can safeguard you from many frauds, malwares, etc.
- iii. It is better to browse using private mode of the browser as well as use https instead of http protocols.
- iv. Whenever downloaded any programs or documents, it is better to download from the genuine source or sites. If downloading from anonymous websites then it's always advisable to scan the file using updated antivirus or scan using sites like virustotal.com. Also it is better to use software like sandoxie and run those downloaded application/documents inside it.
- v. Never allow any body to cyber bullying on you.

- vi. Use of security software like firewall, UTM, etc is always the best way to safeguard from cyber criminal. Use of software like VPN, tor, etc for accessing or communicating over sites which have critical information is always preferred.

2. In case of system security

- i. Operating system should be updated with latest patches available to prevent operating system hacking.
- ii. Application level hacking is possible using the vulnerability of particular application. To remain safe from it's always better to download application from genuine sites & run the downloaded software with the help of application like sandboxie.
- iii. Use of Antivirus for both system as well as online activities is always advisable.
- iv. Never use any usb/cd/dvd devices that you got from any unknown sources or places.

The Need of Action

Internet is now richer by Two Billion global populations. Today, in our day to day living most of the devices we use are integrated. Users, machines, applications and data are in total sync and the rich interaction that internet is providing us. Hence data is getting generated constantly. It has completely re-designed the complete structure of the INTERNET. And not only is that it acting as a growth catalyst for the economy. Social Media, as we have seen in many recent cases can call for a change. And keeping into consideration the Critical Information Infrastructure which includes Banks, Power Plants, Transportation, Intelligence etc. there is an urgent need to keep the cyber space SAFE.. FREE.. and..SECURE. The world has become a small place now. It is merging as a collective Society. The boundaries have vanished, geography has blurred and distance is fading leading to a shared Cyber Space.

Conclusion

The Cyber Security Policy should be dynamic in nature. As technology is developing at a speeding rate, so are these crimes. So the Policy should be tuned to the changing need of the society. It should be designed such that it is long term and short term consistent, realistic and objective based. Constant review to improve the policy and to remove the impudent be made.

There are questions in every mind like:

- I. Who is responsible for ensuring this cyber space to be threat free? Is it Government or Private Sector? or Users ? or Academicians ?

- II. Whether government should take initiative for having Hacker group of the country like US, China, etc or not?
- III. Shouldn't there be some initiative to have an awareness program to safeguard citizen of the country?

Though the governments are taking Cyber Threats Seriously, they need to assure certain basic rights to its users and its citizens and priorities there action and endorse them. It is the governments' duty to protect the fundamental Rights of its citizen in the virtual world i.e. the cyber space but it is the responsibility of person concern to take all security measures to keep themselves safe from the threats of cyber space. Keep surfing & remain alert to remain safe online/offline of this ocean, Internet.

References

1. www.nmap.org/
2. <https://www.wikipedia.org/>
3. <http://thehackernews.com>
4. www.blackhat.com
5. www.sandboxie.com/
6. www.securityweek.com
7. www.cyberforensics.in
8. [www. ncrb.gov.in/](http://www.ncrb.gov.in/)

Cloud Computing Scenario in Indian SMEs: An Empirical Study

*Sanjeev Kumar Singh
Bipul Shyam Purkayastha*

Abstract

As per the NASCOM study the cloud computing is going to have a significant impact on the service industries (mostly SMB's) in the coming year. Another study shows that 19% of the organizations are using the cloud services for production computing and 20% uses for storage for the most of the IT requirements. As per the Economic Times report, 2010, there will be a significant surge in the cloud adoption by Indian SME's. This paper will study the issues and challenges relating the cloud adoption by Indian SME's. A comparative performance and profit analysis of Indian SMEs' (Small and Medium Enterprises) will also be studied who have adopted the cloud services such as (SaaS, PaaS and IaaS) and those who have not adopted.

Keyword: *Cloud Computing, SME, SMB, IaaS, SaaS and PaaS*

Introduction

The cloud computing could be best understood as computing resources (such as application software, operating system and other utility software) kept in a system which is shared by remotely located users, which are connected through the high speed internet. The metaphor "cloud" is used as the system, where the computing resources are kept, is invisible to the remotely located users. The cloud computing can be owned by individual and users can connect to the server and use the computing resources as per requirement. Here the user pays

for the facility he/she has used. This cloud is called as private cloud. A cloud server owned by a company or some agency/organization where the service is open for public use. There is not much difference between private and public cloud service except the cost, security and quality of service. In the public deployment model, the infrastructure is available for general public and owned by the large cloud service provider. The public cloud service provider such as Amazon, Microsoft and Google own and operate the infrastructure at their data center and the end user can access their cloud via internet. Hybrid cloud is a combination of both, public and private. Large cloud service provider can allow the private cloud service provider use their infrastructure and network. This type of cloud used by the end user is cost effective also the private cloud provider will not have to setup their own infrastructure. They pay for the rental of the infrastructure used from the giant cloud service provider.

The type of service provided by the cloud is IaaS (Infrastructure as a Service), PaaS (Platform as a Service) and SaaS (Software as a Service). Clearly the cloud provider has three type services and the end user in this paper is the SME and SMB can request for one or all the services for the service provider.

The SME (Small Scale Enterprise) or SMB (Small Scale Business) are the business entities with low input cost. They use the IT as services to support their business. Automation of different unit of the business can reduce the input cost of the SME or SMB and increase the production, so the profitability. So the cloud service can be one of the options for this SME/SMB.

In this paper our aim is to study the cost and benefit of the SME or SMB if they are using the cloud service. A comparative study will be done using the data.

Cost and Benefit Analysis of Cloud Computing in SME

The benefit of the computing is many. But the benefits depend on the SMB/SME that, what kind of services is sought for. The cost and benefits' of the cloud computing depends largely on the cost of the services sought, cost of the risk management and the agility.

Cost Optimization

Several researches have that the cloud can reduce the cost three forth of the total input cost and thereby increase in the throughput of the company.

Reduce Input Cost

As the SME or SME has many units which needs the automation so that the data can be saved in electronically available. As the business grows the volume of data increases. These data needs to saved and maintained. The cloud computing allows companies to depreciate the cost over three to five years; however, many IT projects do not last three to five years.

Cost Analysis

As the hidden cost is an issue in cloud computing the cloud provider must be able to approximate all the occurring and non-re-occurring cost of cloud computing. A cloud provider should be able to pass on the benefits of the economies-of-scale and experience derived from providing standardized services to multiple clients. The reality is that it sometimes may be less expensive for an organization to build and manage IT resources in-house for certain application workloads that are predictable. It is important to understand the costs for a client to manage their own resources in-house. Key cost components include: server hardware, networking equipment, software licenses, in-house software development and integration, compliance, colocation, utilities, network services, software and hardware support, system administration (headcount), and OS and hardware patching and maintenance

Efficiency

Cloud computing increases the efficiency and simplifies provisioning, de-provisioning and re-deploying resources through automation. The efficiency of cloud computing reduces the amount of time an IT systems administrator has to spend on managing and supporting infrastructure. The average number of server administrators to servers in a typical data center is 50 servers: 1 administrator. The average ratio of cloud-based data centers is 500:1 [5][8].

Pay-per-use

End-users of cloud computing pay only for the resources they use. For example, a cloud end-user may need to use twenty servers to test and develop an application over the course of a few months. Rather than having to buy the hardware, and power to support the temporary project, the company can choose to use only 20 servers for a specific time and pay only for the real time use.

On-Demand Capacity and Elasticity

One of the great concern in using the cloud computing is the real time use of resources. Many a time it may happen, the server is kept in use mode but actually no work is being done. The end-user client must install automated on-use-stop type of mechanism to avoid from paying. Being able to adjust resource usage dynamically and on-demand allows companies to reduce over provisioning.

Optimization of risk

A myth has to be understand in right perspective that cloud computing has the risk of data theft, hacking and data-misuse. Reality is different. Using the cloud allows companies to reduce risk.

Recovery Mechanism

The cloud can be used as a cost-effective way to back-up data or provide a disaster recovery site to improve business operation.

Operational Complexity

The cloud allows companies to reduce operational complexity by simplifying the way in which IT resources are used, managed and delivered.

Data Independency

Data independency often requires an organization restrict the transmission and storage of data to a specific location... Using a public cloud or deploying a private cloud in a country or region allows companies to better meet data sovereignty requirement.

Benchmarking Services

In a large global enterprise, ensuring benchmark is critical to efficiently administering and supporting IT. Because the cloud (SaaS, PaaS, IaaS) delivers a specific set of services, it is easier to manage and reduces errors and support complexity.

Governance, Compliance, Tracking and Control

Governance, compliance tracking and control requirements are the abilities to audit, track, secure and control access to data and company resources. Cloud providers can improve the way in which resources are tracked, secured and controlled by end users, often improving an enterprises ability to be compliant and meet governance requirements.

Skilled Manpower

Organizations often lack the internal skills to migrate to new models and platforms, such as cloud computing. Lacking hands-on experience, internal skills and operational benchmarks, increases the chance of failure, and lengthens the time required for cloud-related initiatives. Relying on an experienced cloud partner reduces risk and the migration from one system to another and any update when the business expands. In this situation experienced cloud partner can add, upgrade the system at a very low cost, thereby increasing output of the company.

Business agility

Agility is the ability of a business to adapt rapidly and cost efficiently in response to changes in the business environment. Strategic agility, or "business agility," can be achieved by quickly adapting goods and services to meet customer demands [3][5][9]. Agility is a concept that incorporates the ideas of flexibility, balance, adaptability, and coordination under one umbrella. In a business context, agility typically refers to the ability of an organization to rapidly adapt to market

and environmental changes in productive and cost-effective ways. In the context of cloud computing, agility often refers to the ability to rapidly develop, test and launch software applications that drive business growth. Cloud computing has the ability to develop the new application, rolling out a new client-facing SaaS application, entering a new market to reach a new client segment, or exiting a market

Achievement of Agility

Market Adaptability

Cloud computing allows companies to significantly decrease the time it takes to provision and de-provision IT infrastructure, speeding delivery of IT projects that are critical to revenue growth or cost reduction. It has the strong adaptability to the market demand

Adaptive Systems

Because of the API accessibility of cloud platforms, it is typically easier to automate IT infrastructure and provisioning in a cloud environment. Integrating business intelligence and analytics platforms and IT monitoring tools with the cloud allows systems to be more adaptive. For example, new servers can be automatically provisioned (or de-provisioned) when load balancing thresholds are met.

Rapid Development

Cloud computing allows companies to significantly decrease the time it takes to provision and de-provision IT infrastructure, speeding delivery of IT projects that are critical to revenue growth or cost reduction.

New Business Models and Innovation

Cloud computing allows companies to better align IT infrastructure and management costs with success and scale as new markets succeed.

Flexibility / Scalability

The ability to rapidly increase or decrease resources on-demand to meet unpredictable application development or production needs. In the case of cloud computing, this might mean being able to spin up 10x to 100x the average utilization of computing resources to support a new project or sudden burst in demand or website traffic. Due to the pay-per-use flexibility of the cloud, end-users are able to scale fast or "fail fast" based on the demands of the business [10] [12].

Leveraging Resource

Multi-tenant architectures (in a private or public cloud) allows users to take advantage of better leverage and scaling cost/benefit for IT resources.

Cloud Computing scenario in India

India can be called as a hub or major player as far as the cloud technology is concerned. As per a report by NASSCOM, cloud computing is said to have a compelling effect on the service industry, which includes the service model, delivery mechanism and business models. The Indian market is expected to cross the mark of \$16 billion by 2020 in the sphere of cloud computing [5]. As per the IDC report the digital data in India will elevate up to 2.3 million petabytes by 2020. An anonymous report, highlights that the private cloud has helped the Indian companies to save up to 50% cost on their infrastructure, which in return increases the number of job opportunities [5]. A survey by Gartner suggests that the Indian companies will adopt the new cloud services at a very fast pace. The CTO's believe that with the improvement in technology and increase in the number of smartphones will lead to the benefit of the entire ecosystem and a sudden rise will be seen in the cloud based services in India. They further suggest that cloud helps to generate automated security management which will ensure the existence of proper and auditable security policies and controls [8] [11]. In short cloud computing has a long way to go and the Indian market will flourish in the coming years and will make most of the cloud technology [2].

Ranging from names like Tata Consultancy Services, Infosys, Wipro and Zenith InfoTech to others like Synapse India, CtrlS, Ozonotel Systems and App Point, every major IT player in the country has jumped on to the cloud bandwagon already. Add to this the commitment that international biggies like IBM and Microsoft are promising the Indian market and the going is not only good, but also great! [6]

At Tata Business Support Services Limited, Business Intelligence in the Cloud environment is simplified.

Use of Cloud Computing (Scenario for Indian SME)

Over the past few years, Indian SMEs have shown immense potential in a variety of industry verticals. They have also shown great interest towards the adoption of technology. Some SME segments have emerged as technology-driven verticals. SMEs in e-commerce have produced great examples by using cutting-edge technologies. Adoption of mobility-based solutions, cloud solutions and the use of hand-held devices have become common among SMEs. Functional areas like marketing and sales are becoming optimized by the usage of technology [8]. In this backdrop, SMEs have become an important focus for Symantec. SMEs are also exploding in the market in terms of growth of data and device growth. Since the dependence on data is continuously growing and the role of technology is becoming decisive on the performance of business, security of data must remain an important aspect of IT infrastructure for every small and medium

business. Cloud solutions are becoming very common among SMEs. Industry stakeholders and even government-level policymakers have understood that cloud computing can become great business enablers for SMEs. The infrastructure has also become favorable for technologies like cloud. Cost and availability of bandwidth have become more appealing, enabling any SME to introduce any new business application on cloud environment. Apart from upfront cost savings, cloud computing also allows entrepreneurs to plan their IT expenditure as an operational expense instead of a capital expense. Solutions like end-point security, e-mail archiving and back-up on cloud are becoming very popular on the cloud environment. Although it is an evolving model, SMEs are aggressively showing an interest on these fronts while adopting cloud solutions [13][14]

Results

The research in this paper is descriptive in nature. The data gathered here was in the form of information collected from the internet posted by the heads of the SME or SMB. As per the data, result shows that the capital expenditure of SME can be reduced by 19% if the company uses the computing as a service though this can vary as per the magnitude of SME. Also, is observed that if the company subscribe the SaaS and IaaS as the cloud computing services, the input cost can be reduced even more than 19% and the output of the company can be increased by 20%.

Conclusion

The study in this paper was descriptive in nature. Due to the time constraint all the data collected through the questionnaire could not be analyzed. Most of the data collected from the internet. The interviews given by the executive head of the SME/SMB were considered as the data for analysis. The result showed the drastic reduction in the capital expenditure and reduced labour cost of the company, if the company chooses to use the cloud computing. It was also observed that most of the SME/SMB used IaaS and SaaS as a cloud computing services. The future study on the cloud computing can be the real time comparison of cost-benefit analysis.

References

1. B.B Aggarwal and M. Barnes, "The Case for Cloud CRM in India", White Paper, Springboard Research, 2010.
2. <http://www.gartner.com/resId=2174917>.
3. A. Kambil, "A head in the clouds", vol. 30, no. 4, pp. 58-59, 2009.
4. R. Fox, "Digital Libraries: The systems analysis perspective", Library in the Clouds, vol. 25, no. pp. 156-161, 2009.
5. http://www.dnb.co.in/smes/company_listing.asp

6. Cloud Computing -Wikipedia.
7. Monika Sharma, Ashwani Mehra, Haresh Jola, Anand Kumar, Dr. Madhvendra Misra, and Ms. Vijayshri Tiwari "Scope of cloud computing for SMEs in India", *Journal of Computing*, Volume 2, Issue 5, May 2010, ISSN 2151-9617.
8. <http://www.bisinfotech.com/article/cloud-computing-the-indian-scenario/>
9. N. Leavitt, "Is cloud computing really ready for prime time?" vol. 42, no.1, pp 15-20, 2009.
10. D.C. Wyld, "The utility of Cloud Computing as a new pricing-and consumption-model for information technology", Vol. 1 No.1, 2009.
11. <http://www.tata-bss.com/in/future-of-cloud-services-in-india-and-trends/>
12. K. Hartig, "what is cloud computing?", 2008.
13. <http://www.supportbiz.com/articles/tech-mate/cloud-becoming-major-enabler-among-smes.html>.
14. W. Ashford, "Cloud clears the way for SME innovation (cloud computing)", 2008.

Proposing a Solution to Dynamic Mac Address Change of a Virtual Machine

Virtual Machine is installed on a Host Machine
Using a Type- II Hypervisor

Partha Pratim Das
Rini Elis Tirkey
Jyotismita Talukdar
Ankit Singh

Abstract

A virtual machine is a software implementation of a computing environment on which an operating system or a program can be installed and run. It is hosted by a physical computing environment so that it can use the resources like CPU, hard disk and other hardware resources which are managed by a virtualization layer. Multiple VM environment can be created on a single physical host OS, which are isolated from each other. Each virtual machine has its own IP address assigned by the DHCP server from a pool of IP addresses. Typically a VM's initial MAC address is defined in the configuration file (.vmx). On re-powering the VM, the MAC address changes dynamically. This can cause problems like faulty MAC filtering in a network. We provide a helping hand in solving the issue.

Key Words: *Hypervisors, DHCP, MAC Address, IP Address, SUSE Linux Enterprise Server, VMware Workstation, Bridged Network, Spanning, Tree Algorithm.*

Introduction

Hypervisors

In computing, a hypervisor or virtual machine monitor (VMM) is a piece of computer software, firmware or hardware that creates and runs virtual machines. A computer on which a hypervisor is running one or more virtual machines is defined as a host machine. Each virtual machine is called a guest machine. The hypervisor presents the guest operating systems with a virtual operating platform and manages the execution of the guest operating systems. Multiple instances of a variety of operating systems may share the virtualized hardware resources. Types of Hypervisors are:

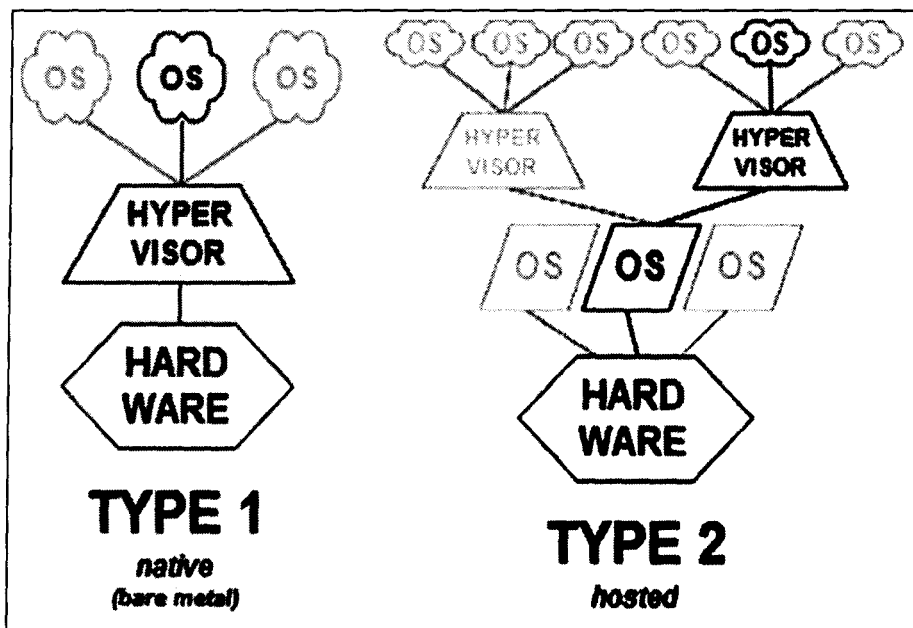


Fig. 10.1: Types of Hypervisors

Type 1 (or native, bare metal) hypervisors run directly on the host's hardware to control the hardware and to manage guest operating systems. A guest operating-system thus runs on another level above the hypervisor. Modern equivalents include Oracle VM Server for SPARC, Oracle VM Server for x86, the Citrix Xen Server and VMware ESX/ESXi.

Type2 (or hosted) hypervisors run within a conventional operating-system environment. With the hypervisor layer as a distinct second software level, guest operating-systems run at the third level above the hardware. For example VMware Workstation and Virtual Box.

VMware Workstation

VMware Workstation is a hypervisor that runs on x64 computers; it enables users to set up multiple virtual machines (VMs) and use them simultaneously along with the actual machine. Each virtual machine can execute its own operating system, such as Microsoft Windows, Linux or BSD variants. As such, VMware Workstation allows one physical machine to run multiple operating systems simultaneously.

VMware Workstation supports bridging existing host network adapters and share physical disk drives and USB devices with a virtual machine. In addition, it can simulate disk drives. VMware Workstation can save the state of a virtual machine in one point of time. These saved states, known as a "snapshots", can later be restored, effectively returning the virtual machine to the saved state.

VMware Workstation includes the ability to designate multiple virtual machines as a team which can then be powered on, powered off, suspended or resume as a single object, making it particularly useful for testing client-server environments.

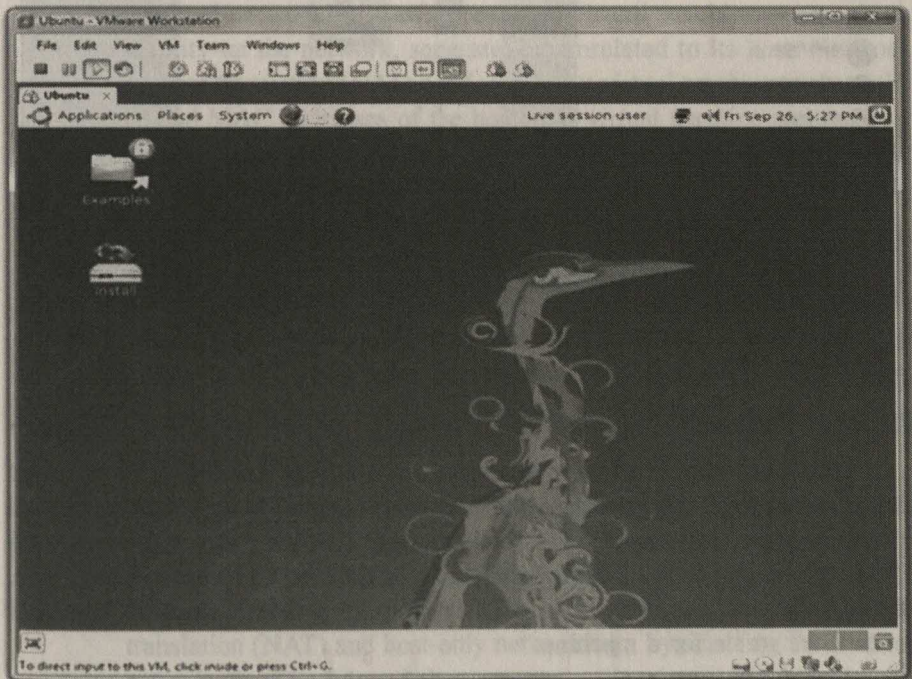


Fig. 10.2: VMware Workstation

SUSE Linux Enterprise Server

SUSE Linux Enterprise Server (SLES) is a Linux-based operating system developed by SUSE. It is designed for servers, mainframes, and workstations but can be installed on desktop computers for testing as well.

The current version is SLES 11 SP3, released July, 01 2013, which is developed from a common codebase with SUSE Linux Enterprise Desktop and other SUSE Linux Enterprise products.

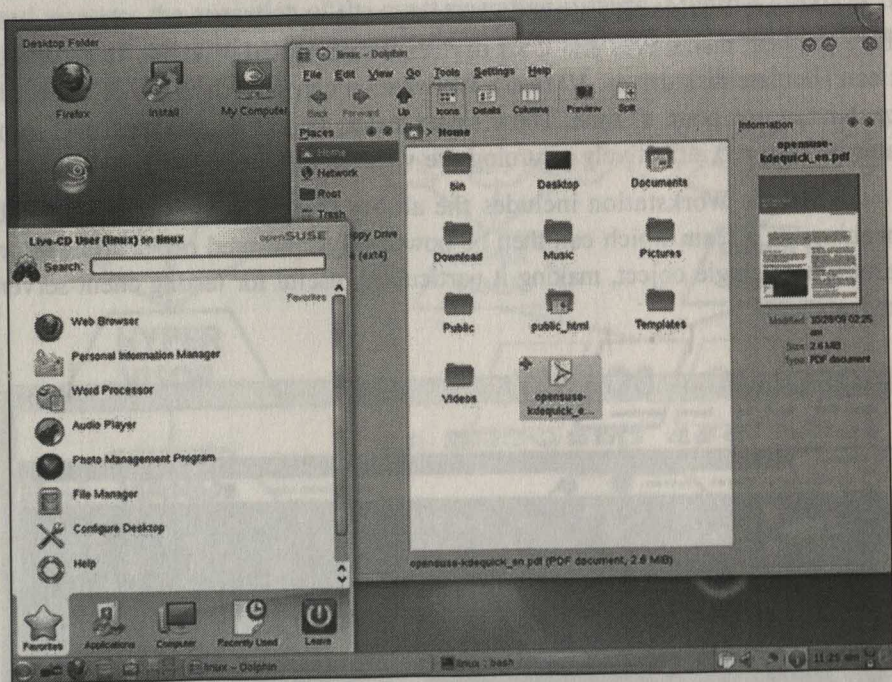


Fig. 10.3: SUSE Linux Enterprise Desktop

Problem Statement

The MAC address of the net adapters of a virtual machine is regenerated when we repeatedly plug the network (Ethernet) cable of a workstation installed with hypervisor, instead of prescribed settings. This problem was tested with VMware Workstation 9.0.2 being installed on SUSE Linux Enterprise Server SP2. Due to MAC address change, user faces multiple problems, as MAC address needs to be stable for a machine.

Alternate problem : On being connected to a network two virtual machines might get a MAC address conflict, which will eventually lead to IP allocation rejection by DHCP.

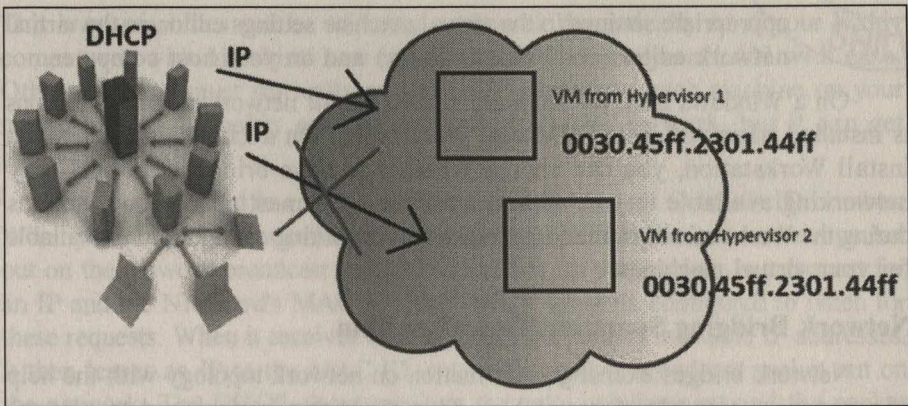


Fig. 10.4: DHCP

VMWARE Network Connections

VMware Workstation provides several ways you can configure a virtual machine for virtual networking:

- A. Bridged networking configures your virtual machine as a unique identity on the network, separate and unrelated to its host. Network address translation (NAT) configures your virtual machine to share the IP and MAC addresses of the host. The virtual machine and the host share a single network identity that is not visible outside the network. NAT can be useful when one is allowed a single IP address or MAC address by one's network administrator.
- B. Host-only networking configures your virtual machine to allow network access only to the host. This can be useful when you want a secure virtual machine that is connected to the host network, but available only through the host machine. See Host-Only Networking.
- C. Custom networking lets one configure your virtual machine's network connection manually.
- D. If one select the typical setup path in the New Virtual Machine Wizard when you create a virtual machine, the wizard sets up bridged networking for the virtual machine. You can choose any of the common configurations - bridged networking, network address translation (NAT) and host-only networking - by selecting the Custom setup path. The wizard then connects the virtual machine to the appropriate virtual network.
- E. You can set up more specialized configurations by choosing the

appropriate settings in the virtual machine settings editor, in the virtual network editor (on Windows hosts) and on your host computer.

On a Windows host, the software needed for all networking configurations is installed when you install VMware Workstation. On a Linux host, when you install Workstation, you can choose whether to have bridged and host-only networking available to your virtual machines: you must choose both options during the Workstation installation to make all networking configurations available for your virtual machines.

Network Bridging Spanning Tree Algorithm

Network bridges exchange information on network topology with the help of a spanning tree algorithm (STA). This algorithm is designed to prevent forwarding topology loops in networks with redundantly bridged segments. In a network with topology loops, bridges can cause forwarding storms by relaying the same frame endlessly from one segment to another.

The STA allows a bridge to select the most efficient path when faced with multiple paths in a multi-segment network. It provides each bridge with the information required to disable one or more of its ports to ensure that the network topology is loop-free. The STA can easily adapt to topology changes. It uses timers to ensure that the forwarding topology is recomputed if a bridge is removed or shut down.

In a multi-segmented network, each bridge must have a globally unique identification number. Most implementations take advantage of the global uniqueness of Ethernet MAC addresses and use the lowest-numbered MAC address from among its adaptors as the bridge identification number.

Network bridges communicate with each other through special packets, which can be referred to as STA packets. STA packets submit information about the network forwarding topology from one bridge to another. Each STA packet contains the following pieces of information:

- A. The identifier of the bridge that the transmitting bridge assumes to be the root bridge.
- B. The cost (number of hops) from the transmitting bridge to the bridge it believes to be the root bridge.
- C. The identifier of the transmitting bridge.

DHCP-Assigning IP Address

DHCP, Dynamic Host Control Protocol, is a method for assigning IP addresses to hosts on a network. A DHCP server running on a network manages a pool of available IP addresses and doles them out to hosts that request them.

This way, machines can be added to a network dynamically so if your buddy comes over with his shiny new laptop, he can hop on your LAN with ease. Otherwise, you must statically assign an IP address to each machine on your network. This probably easy to manage in a home network, but it can get annoying quickly

Every NIC card has a unique MAC (Media Access Control) ID number. When you plug a host into the network running a DHCP client, it sends a packet out on the network broadcast (255.255.255.255). In that packet, is a request for an IP and the NIC card's MAC ID. The DHCP server is configured to listen for these requests. When it receives one, it checks the pool of available IP addresses. It then leases an IP to the MAC ID, returning another broadcast packet out on the network. The DHCP client receives the acknowledgement and the packet contains data telling the host it's IP and generally your other routing information. The DHCP reply packets will usually handle the job of telling a host it's default route, it's DNS servers, and many other configurable pieces of information as well. Below is a sample of a dhcpd.conf file:

```
# $OpenBSD: dhcpd.conf,v 1.1
```

```
1998/08/19 04:25:45 form Exp $
```

```
#
```

```
# DHCP server options.
```

```
# See dhcpd.conf(5) and dhcpd(8) for more  
information.
```

```
#
```

```
# Network:
```

```
192.168.1.0/255.255.255.0
```

```
# Domain name:
```

```
my.domain
```

```
# Name servers:
```

```
192.168.1.3 and 192.168.1.5
```

```
# Default router: 192.168.1.1
```

```
# Addresses:
```

```
192.168.1.32 -
```

```

192.168.1.127
#
shared -network LOCAL -NET {
    option                                domain -name
                                "my.domain";

option domain -name -servers 192.168.1.3, 192.168.1.5;
subnet 192.168.1.0 netmask 255.255.255.0 {
option    routers 192.168.1.1;
                                range 192.168.1.32 192.168.1.127;

```

Fig. 10.4: Sample of a dhcpd.conf File

It sets up a pool of dynamically assignable IPs from 192.168.1.32 to 192.168.1.127 inside of network 192.168.1.0/24. Furthermore, it tells those machines that their default route is at 192.168.1.1, they have a domain name of "my.domain" and name servers at 192.168.1.3 and 192.168.1.5. The sample file reserves the first 30 IPs in 192.168.1.0/24 for static IP machines.

The only caveat with DHCP is that you must have your network carefully segmented. You should not have two DHCP servers attempting to assign IPs on the same network segment. So if you have a DSL connection to the Internet, you need to have two IPs in one Linux box that connects to both the DSL router and the other to a hub connected to your internal LAN. The reason is that your Internet provider will also be attempting to assign IPs via DHCP and you will have problems if your network is not segmented this way.

```

2013-06-03T05:28:42.554-05:00 vmc N28: _____
2013-06-03T05:28:42.554-05:00 vmc N28: PREF Failed to load user preferences.
2013-06-03T05:28:42.572-05:00 vmc N28: Hostname=arc-1609
2013-06-03T05:28:42.572-05:00 vmc N28: P=127.0.0.1 (b)
2013-06-03T05:28:42.572-05:00 vmc N28: P=127.0.0.2 (b)
2013-06-03T05:28:42.572-05:00 vmc N28: P=10.40.11.125 (nll)

```

2013-06-03T05:28:42.5720500	vmx	H20:	P=172.16.219.1	(vmmnet1)
2013-06-03T05:28:42.5720500	vmx	H20:	P=192.168.3.1	(vmmnet0)
2013-06-03T05:28:42.5720500	vmx	H20:	System uptime	102570000 os
2013-06-03T05:28:42.5720500	vmx	H20:	Command line:	AutoRun\vmware\bin\vmware-vmx #NAME? vmx.noLibBuildNum

Following Log Files were recorded:

Assigning MAC Address

Each host has a pool of MAC Addresses. By default the Pool holds 256 Addresses, and when asked gives them out in a Round Robin fashion. The MAC address is stored in the machines XML configuration file.

So the MAC address has 3 bytes fixed "00-15-5d", followed by 2 bytes specific to the current host, and one byte as the individual MAC in this Pool. This algorithm works fine on a standalone host until you create less than 256 NIC's. Due to the Round Robin method, addresses are given out again.

Every time a VM is started, the dynamic MAC address assigned to the NIC is checked against this pool. If the MAC is outside of this pool, a new MAC is taken from the Pool and assigned to the NIC. Again, this check is done during VM startup only. When a virtual machine is powered on, VMware Workstation automatically assigns each of its virtual network adapters an Ethernet MAC address. The virtual machine is assigned the same MAC address every time it is powered on, so long as the virtual machine is not moved (the path and filename for the virtual machine's configuration file must remain the same) and no changes are made to certain settings in the configuration file.

ARP

Short for AddressResolutionProtocol, a network layer protocol used to convert an IP address into a physical address (called aDLC address), such as an Ethernet address. A host wishing to obtain a physical address broadcasts an ARP request onto the TCP/IP network. The host on the network that has the IP address in the request then replies with its physical hardware address.

There is also Reverse ARP (RARP) which can be used by a host to discover its IP address. In this case, the host broadcasts its physical address and a RARP server replies with the host's IP address.

```
C:\Users\N0lan>arp -a
```

Interface: 10.10.13.113 --- Wsh		
Internet Address	Physical Address	Type
10.10.13.1	00-0c-29-ee-55-7b	dynamic
10.10.13.110	00-11-ee-00-ff-ee	dynamic
10.10.13.118	00-22-5f-05-0c-3c	dynamic
10.10.13.121	00-0c-29-ee-55-7b	dynamic
10.10.13.255	ff-ff-ff-ff-ff-ff	static
224.0.0.22	01-00-5e-00-00-16	static
224.0.0.251	01-00-5e-00-00-1b	static
224.0.0.252	01-00-5e-00-00-1c	static
239.255.255.250	01-00-5e-7f-ff-fa	static
255.255.255.255	ff-ff-ff-ff-ff-ff	static

Interface: 192.168.150.1 --- Wsh		
Internet Address	Physical Address	Type
192.168.150.255	ff-ff-ff-ff-ff-ff	static
224.0.0.22	01-00-5e-00-00-16	static
224.0.0.251	01-00-5e-00-00-1b	static
224.0.0.252	01-00-5e-00-00-1c	static
239.255.255.250	01-00-5e-7f-ff-fa	static

Fig. 10.5: Reverse ARP (RARP)

Proposed Solution

- Storing the previously allocated IP address in Config.xml

Every VM has a Config.xml file which contains information like- State, Time, Drivers, I/O Devices, etc. We would propose to add another parameter which would store the last allocated IP address of the virtual machine in a network.

- Inside DHCP server and Hypervisor

We suggest of creating a .xml file like ARP table to be stored in the DHCP server and the Hypervisor, which will hold the previously assigned MAC addresses of the VMs along with the IP addresses.

Rule: The IP column values cannot be duplicated

Firstly when a VM is started or requested for in a hypervisor, the table will be verified at first- the existing IP will be checked and if it exists the corresponding MAC address will be assigned, otherwise MAC will be randomly generated.

Secondly, when there is a MAC address duplicity of any two virtual machines, separate IP addresses will be given out by the DHCP server, as IP column cannot be generated.

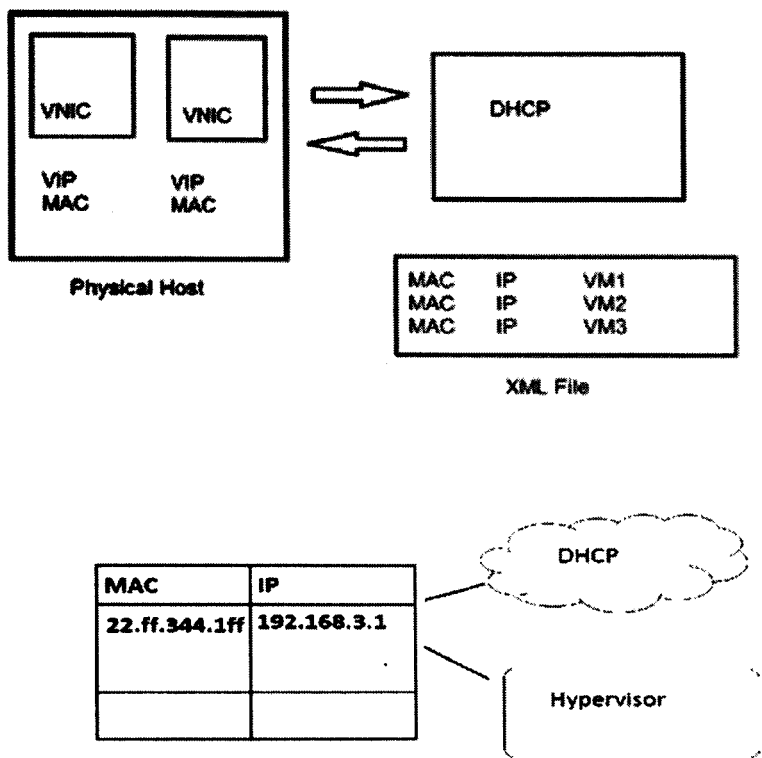


Fig. 10.6: Separate IP MAC Address Duplicity of any two virtual machines

Acknowledgement

We would like to express our heartiest gratitude to the faculties of University of Technology and Management, Shillong specially Mr. Deepak Shukla (Senior Assistant Professor). We would like to thank them for their guidance and support for this paper.

References

1. <http://en.wikipedia.org/wiki/Hypervisor>
2. http://en.wikipedia.org/wiki/VMware_Workstation
3. http://en.wikipedia.org/wiki/SUSE_Linux_Enterprise_Server
4. <http://blogs.msdn.com/b/robertvi/archive/2011/03/25/how-does-a-nic-inside-a-vm-get-a-mac-address-assigned.aspx>
5. <http://msdn.microsoft.com/en-us/library/ms883143.aspx>
6. <http://www.webopedia.com/TERM/A/ARP.html>

Categorization of Traffic Hijacking Techniques in Cloud Ecosystem

Narayan Tripathi

Pallavi Yadav

Prateek Chaturvedi

Abstract

Cloud Computing is a way to increase the capacity or add capabilities dynamically without investing in new infrastructure, training new personnel, or licensing new software. It extends information technology's existing capabilities. In last few years, cloud computing has grown from being a promising business concept to one of the fast growing segment of the IT industry. But as many enterprises and individual are shifting to the cloud, the amount of information placed on cloud is increasing on the day to day basis, concerns are arising how secure this environment is. Cloud security is one of the major threats for the development of cloud computing and data privacy and protection are major issues hindering the adoption of cloud computing. One of the major issues for cloud security is traffic hijacking. This paper tries bringing together different traffic hijacking technique in traditional as well as cloud environment and also explains traffic and hijacking in sufficient details so readers could have the idea of what really we are trying to say through this paper and what are the impacts of traffic hijacking in near future.

Key words: *Traffic, hijacking, Traffic Hijacking, Cloud Computing.*

Introduction

Cloud computing is a system, where resources of datacenter are used to their maximum efficiency, reducing the cost and also the infrastructure. The resources are shared using the virtualization technology which also gives a flexible usage to the customers "use as required". Traffic hijacking is a very huge threat over cloud computing. Traffic means flow of data and hijacking means taking control of something and run according to own wish in place of required flow.

Traffic Hijacking is a very huge threat not only for cloud service offerings but for the whole system called Internet. The problem arose when internet came into existence. Traffic hijacking was a problem for the scientists from decades. Many solutions and standards were and are proposed or tried to resolve the problem but till today there is no proper solution which could completely resolve the issue.

Background

A. Traffic

Traffic means the flow of data across the internet.

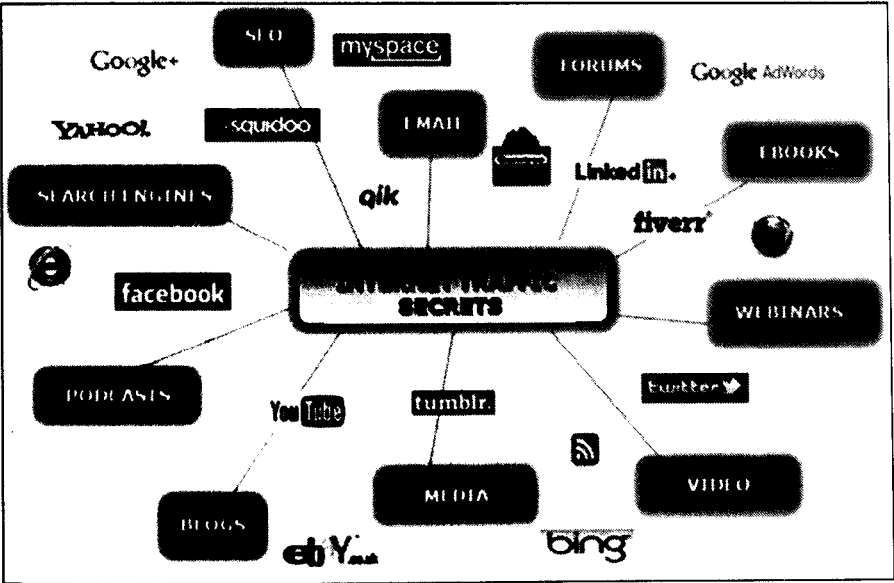


Fig. 11.1: Illustring different Traffics

B. Types of Traffic

- **Busy Traffic**-The pattern of data transmission is uneven.
For example- HTTP, FTP, downloads graphic, video content.

- **Interactive Traffic-** It is a -session; consist of comparatively short request and the response pair. Applications involving real- time interaction with the end user.

For example-web browsing ,Online purchasing, SSL transaction, IM, Telnet sessions.

- **Latency Sensitive Traffic-** Operator has an expectation to deliver on time and in terms we could say that it is a hand to hand transmission. Unlike busy traffic, packets are transferred, generated at particular interval of time.

For example- VoIP, online gaming, video conferencing.

- **Non- Real time traffic-** Application where timely delivery does not matter. Internet protocol like news (NNTP) and sent mail.

For example- E- mail, Batch processing Applications.

- **Background Traffic-** Pertains activities which should not impact the use of network by other users.

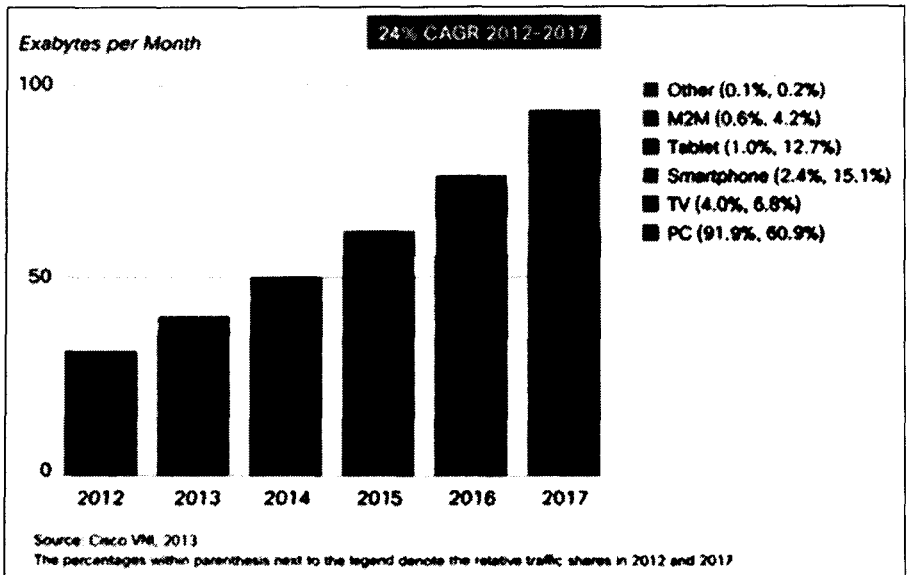


Fig. 11.2: CAGR Survey Report on Internet Traffic

C. Traffic Protocols

LDAP (Light weighted Accesses Protocol): The protocol accessing data from the directory services like: Microsoft directory access.

RADIUS (remote authentication dial in user service): Is a network protocol that provides centralized authentication, authorization, accounting management that connects and use network services.

RDP (Remote desktop protocol): developed by Microsoft, which provide with a graphical interface to connect to another computer across network connection?

DNS traffic: Hierarchical distributed naming system for computer, services or any resources connected to internet or private network.

Certification revocation List: It is also known as online certificate status Protocol (OCSP) is an internet protocol used for obtaining revocation status.

Kerberos: computer network authentication protocol which works on the basis of 'tickets' to allow nodes communicating over non- secure network to prove their identity.

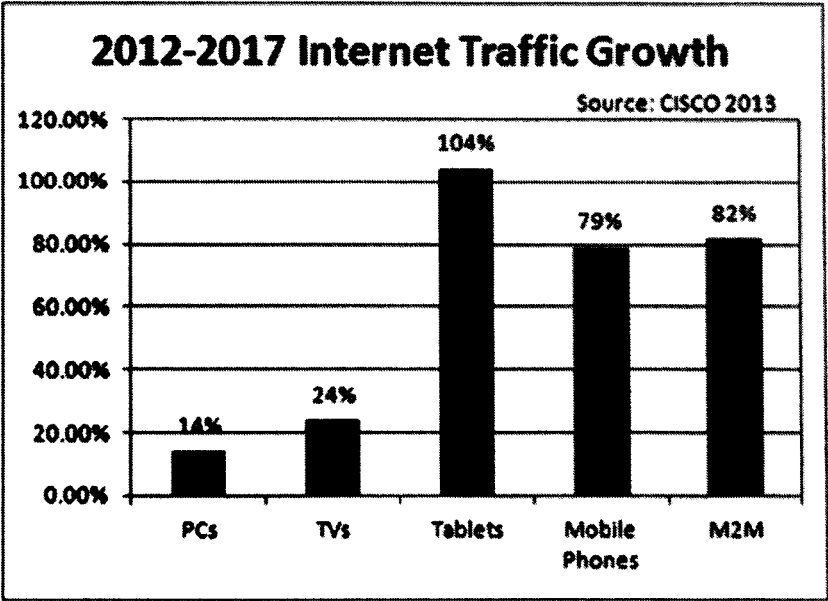


Fig. 11.3 Growth of Traffic

D. Hijacking

It is a type of network security attack which takes control of a communication.

Types of Hijacking

- 1. **Man-in-middle**: In this attack, attacker takes control of an established connection while it is in process.

Attacker's intercepts messages in a public key exchange and then submitting their own public key for the requested one, so that two original parties still appear to be communicating with each other directly.

Attackers use a program that appears to be the server to the client and appears to be the client to the server.

Attack may be used to simply gain excess to the messages, or to enable the attackers to modify them before retransmitting them.

- II. Website hijack: Attackers simply register a domain name similar enough to a legitimate one that users likely to type it, either by mistaking original name or through a type.

Uses

This type of hijacking is currently being employed to send many un-ways users to a pornographic site instead of site they required.

III. Browser Hijacking

While browsing the web or accessing online content, control is displayed, how the browser operators and how it is configured could be changed remotely- this is browser hijacking. For this type of hijacking a type of malware is used that alters the computer's browser setting so that users are directed to website that one has no intention of visiting. Using browser hijackers the person responsible for the spyware can deliver pop-up ads, reset the browser homepage, or direct the browser to websites the victim would not normally visit.

IV. DNS hijacking:

It is also called as DNS redirection. Unauthorized modification of a DNS server or change of DNS address that redirects users attempting to access a web-page to a different web page that looks the same, but contain extra content such as advertisement is a competitor page , a malware page or third - party search page.

Can be obtained in two ways:

- Manipulation by ISP's
- Manipulation by Registrars

V. Domain Hijacking

It is also called domain theft act of changing the registration of domain name without the permission of its original registrant.

VI. IP Hijacking

It is also called BGP hijacking, prefix hijacking or route hijacking.

It is an illegitimate takeover of group of IP addresses by corrupting the internet routing tables.

VII. Page Hijacking

Form of search engine index spamming achieved by creating a rogue copy of a popular website which shows contents similar to a web-crawler, but redirects web surfers to separate or malicious websites. Or in other words we can define it as:

It is a form of clocking, made possible because some web- crawlers detect duplicate while indexing web- pages.

VIII. Reverse-Domain Name Hijacking

(Or reverse cyber-squatting) occurs where trademark owner attempts to secure domain name by making false cybersquatting claims against domain name rightful owner.

IX. Session Hijacking

(Also called cookie hijacking) exploitation of valid computer sessions - sometimes also called a session key. The main motive of this kind of attack is to gain unauthorized access to information or services in computer system.

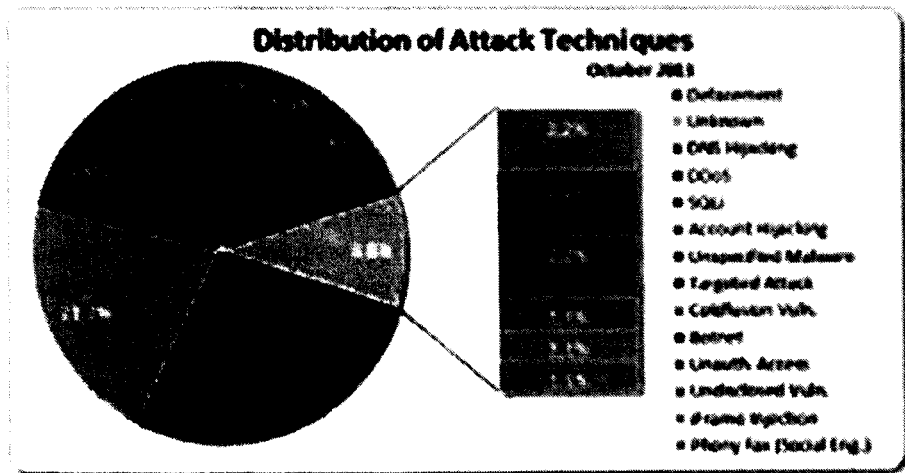


Fig. 11.4: Distribution of Attack Techniques in Cloud Computing

E. Cloud Computing

Definition

It is a type of computing that relies on sharing computing resources rather than having local servers or personal devices to handle applications. In other words it is a next stage in the internet's evolution, providing means through which everything- from computing power to computer infrastructure applications,

business process to personal collaboration - can be delivered to a receiver as a session wherever and whenever you need.

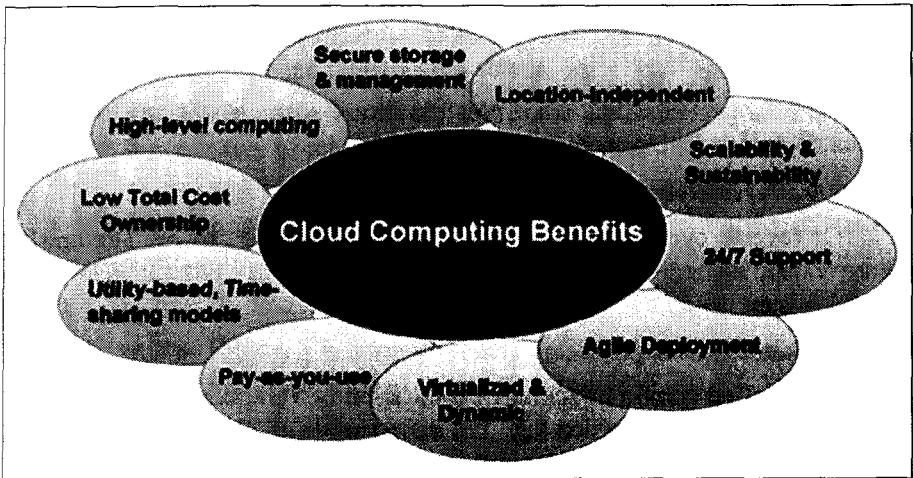


Fig. 11.5: BENEFITS of Cloud Computing

F. Key Differences between Traditional IT and Cloud Computing

- Active agents vs. passive self-service usage of resources.
- Direction of travel and sequence of events.
- Standardized shared infrastructure/ environment.
- Difference in types and degree of user control.

Note: According to cloud security alliance 2013 report, service traffic hijacking is the third greatest cloud computing security risk.

G. Service and Account Traffic Hijacking

In this type of service breach, hackers seek to hijack other's account by stealing his/her security credentials and then caves dropping on the activities and transactions of the person. Hackers can also manipulate the person's data, insert false information and redirect the clients to illegitimate sites.

This form of hijacking comes under the first type of hijacking (Man-in-middle attack).

H. Reasons to Consider

It is scary because hackers are able to use person's reputation and trust he/she has built up to manipulate his/her clients. As we all are aware of, that reputation always matters at any place and anytime, under this situation reputation, trust and confidentiality of a person becomes suspect.

I. Side Effect

When an attacker gains access to credentials another person, they can eavesdrop on the activities and transactions of the victim, manipulate data, return falsified information, and redirect the clients of the victim to illegitimate sites. The account or service instances of the victim may also become a new base for attacker. From here, they may leverage the power of reputation of the person to launch subsequent attack.

J. Range of Attack

These attack range from man-in-middle attack, to phishing and spam complains, to denial of service attack.

K. Suggested Remedies by CSA to Lessen the Threats

- Prohibit sharing of account credentials between users and services.
- Leverage storing two-factor authentication techniques where possible.
- Employ proactive monitoring to detect unauthorized activity.
- Understand cloud-providers security policies and SLA's.

Conclusion

Today, traffic hijacking is considered to be a very important threat and it is not a threat over cloud computing only, it is a threat over internet which IT is facing from a very long time and till today there is not a complete solution for the hijacking over internet. In this paper we have specified the service and account hijacking as we all know the cloud computing has reduced the overheads of the industries. This has given new light to this increasing demand of IT industries but as has been emphasised yet because of the fear of security and leakage of confidential and private data, industries are not completely showing their interest. So, in short this is a fruitful technology in all aspects but needs to be more secure and promising since security and accountability are very important stems of tree called cloud.

References

1. <http://www.slideshare.net/SrinivasDabbeeru/traffic-types-in-internet->
2. <http://mmnetworks.stanford.edu/new/papers/traffictypes.pdf>
3. http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/VNI_Hyperconnectivity_WP.html
4. http://www.scorpioconnect.com/internet_traffic_secrets.html
5. <http://www.websiteoptimization.com/bw/1305/>
6. <http://searchsecurity.techtarget.com/definition/browser-hijacker>
7. <http://www.brighthub.com/computing/smb-security/articles/28384.aspx>

8. <http://www.computerhope.com/jargon/d/dnshijac.htm>http://en.wikipedia.org/wiki/DNS_hijacking
9. http://en.wikipedia.org/wiki/Domain_hijacking
10. http://en.wikipedia.org/wiki/IP_hijacking
11. http://en.wikipedia.org/wiki/Page_hijacking
12. http://en.wikipedia.org/wiki/Session_hijacking
13. <http://en.wikipedia.org/wiki/Hijacking>
14. <http://searchsecurity.techtarget.com/definition/hijacking-hijacking>
15. http://www.webopedia.com/TERM/C/cloud_computing.html
16. <http://www.dummies.com/how-to/content/what-is-cloud-computing.html>
17. <http://ttajts0.tripod.com/cloud/hijacking.htm>
18. <http://cmsreport.com/articles/how-safe-is-your-cloud-data-from-service-traffic-hijacking--5653>
19. <http://www.scl.org/site.aspx?i=ed28054M>.

Networking Aspect in Cloud Computing

*Bharathi Paleti
Vishesh Shukla
UttkarshUpmanyu*

Abstract

Cloud computing relatively a recent term builds on decades of research in virtualization, distributed computing, utility computing and more recently networking, web and software services. It implies a service-oriented architecture, reduced information technology overhead for the end user, great flexibility, reduced total cost of ownership, on-demand services and many other things. The network, similar to CPU and memory, is a critical resource in the cloud. The network serves as the linkage b/w the end users (consuming cloud services) and the provider's data centers providing the cloud services. This paper focuses on how do network architectures affect cloud computing. It will assess the network's role in security, reliability, performance and scalability of cloud computing. Also, the process of evolving the network architecture will be discuss which will provide better support to cloud computing and cloud based service delivery. As the cloud continues to assume a more mainstream role, its expansion is likely lead to new networking demands. Literature survey proves that the cloud meteoric rise is changing data center network requirements. We have to agree that the network will play a major role in guiding the cloud moving forward, whether it grows through a customer centric or vendor focused approach.

Key words: *Cloud computing, Networking.*

Introduction

Cloud computing is an emerging area that affects IT infrastructure, network services, and applications. Cloud computing is a technique of resource sharing where servers and storage in multiple locations are connected by networks to create a pool of resources. When applications are run, resources are allocated from this pool and connected to the user as needed. There are several characteristics of cloud computing such as elasticity and scalability, resiliency, on demand service, multi-tenancy and workload movement. Cloud computing involves different delivery as well as deployment models and based on these models different networking aspects came into picture.

The missions of connecting the resources (servers and storage) into a resource pool and then connecting users to the correct resources create the network's mission in cloud computing. For many cloud computing applications, network performance will be the key to cloud computing performance. Today networking challenges are lagging behind to the computational aspects of cloud.

Cloud data centres (DCs) are built by interconnecting numerous cloud computing elements. Interconnections are performed at two levels: i) among computing elements in the same data centre, and ii) among data centres. In general, all the computing elements and their connections - both intra and Interconnections - are referred to as cloud networks (CNs). Cloud computing poses several challenges due to the complexity of interconnection networks, large number of users, user's mobility, and a large variety of application services.

The key challenges include the in cast problem, cost effective data centre scalability, secured cloud access, unpredictable track patterns and variable demand, dynamic network resource allocation, workload and IP mobility to name a few. This paper takes will takes note of most of these networking aspects of cloud computing, thus leading to some new mechanisms, protocol techniques or the network architecture related to different deployment models of the cloud.

Network Architecture

A. Public Cloud

A public cloud is built over the Internet, which can be accessed by any user who has paid for the service. Public clouds are owned by service providers. They are accessed by subscription. To access a public cloud all one need is to have a machine connected to the internet. Public cloud delivery models will commonly deliver services to support the interests of a broad population.

For cost efficiency, the Internet is used as the basic networking platform for users to connect to the cloud. When employees use public cloud delivery models, organizations are expanding the company security boundary to the

Internet and beyond. Cloud service providers may offer a broad range of available access methods and connectivity technologies (including broadband, wireless and mobile technologies) that allow the cloud provider's services to be accessed anywhere, anytime. Public cloud architecture holds many VM instances running on a single per multiple machines, to provide services to the end users. But it is to be noted that virtualization is not mandatory for cloud. Public cloud uses networking components and technologies to provide its services to the users. Those components include routers , switches , hubs , Ethernet cables , access points etc to provide communication link between intra components as well as between cloud provider data center and cloud user . The architecture includes technologies such as routing, switching etc.

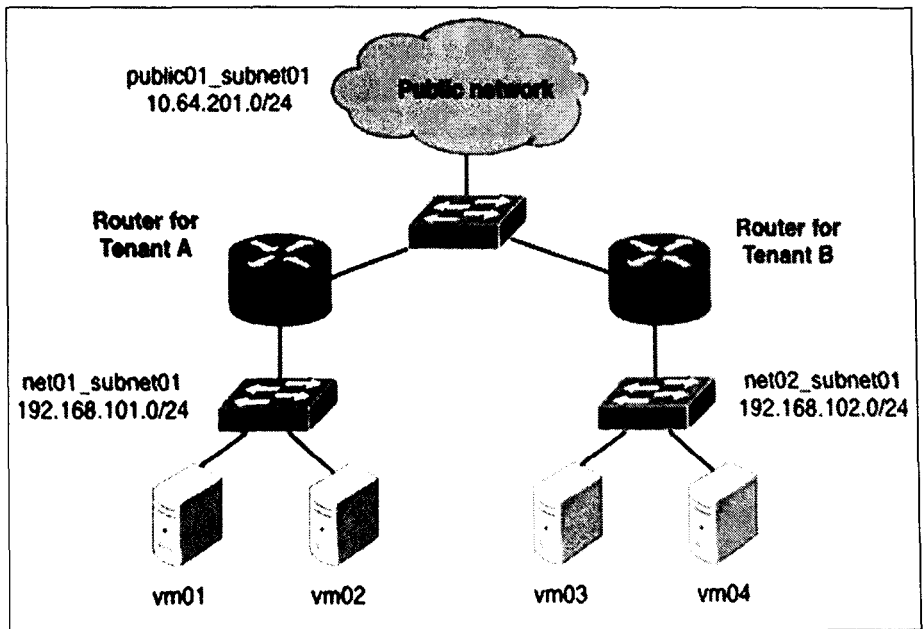


Fig. 12.1: Network Architecture of Public Cloud

B. Private Cloud

The private cloud is built within the domain of an intranet owned by a single organization. Therefore, they are client owned and managed. Their access is limited to the owning clients and their partners. Their deployment was not meant to sell capacity over the Internet through publicly accessible interfaces. Private clouds give local users a flexible and agile private infrastructure to run service workloads within their administrative domains.

Most often, enterprises will access their own private clouds using the same technology they employed for access to their data centers. This may include an Internet VPN or VPN service from a network operator. A VPN is a network that uses a public communication infrastructure, such as the Internet, to provide remote offices or individual users with secure access to their organization's network. A VPN ensures privacy through security procedures and tunneling protocols such as the Layer Two Tunneling Protocol (L2TP). Data is encrypted at the sending end and decrypted at the receiving end. The data is encrypted by using various encryption algorithm such as MD5. These algorithms distort the actual message and convert it to something non-understandable so that data isn't intercepted in transit.

If application access was satisfactory in a "pre-cloud" configuration, a transition to private cloud computing is not likely to impact access performance. Apart from VPN there are several other components such as load balancers, firewalls and several types of encryptions are used with certain technologies.

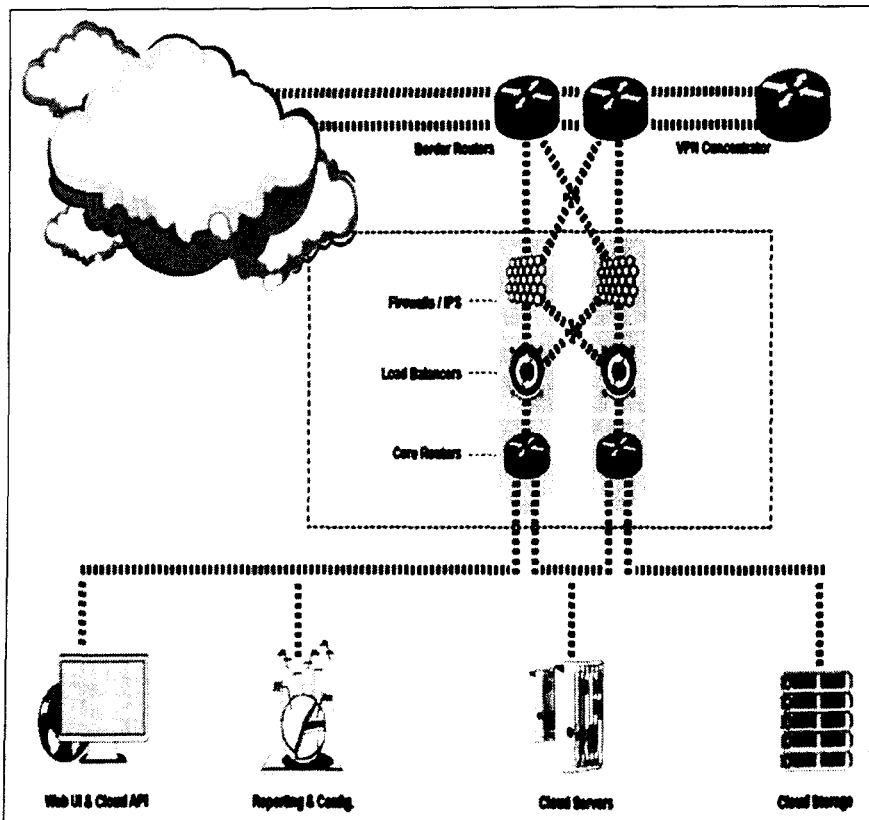


Fig. 12.2: Network Architecture of Private Cloud

C. Hybrid Cloud

A hybrid cloud is built with both public and private clouds. Private clouds can also support a hybrid cloud model by supplementing local infrastructure with computing capacity from an external public cloud. A hybrid clouds provides access to client, partner network, and third party. The hybrid clouds operates in the middle way with public and private cloud advantages.

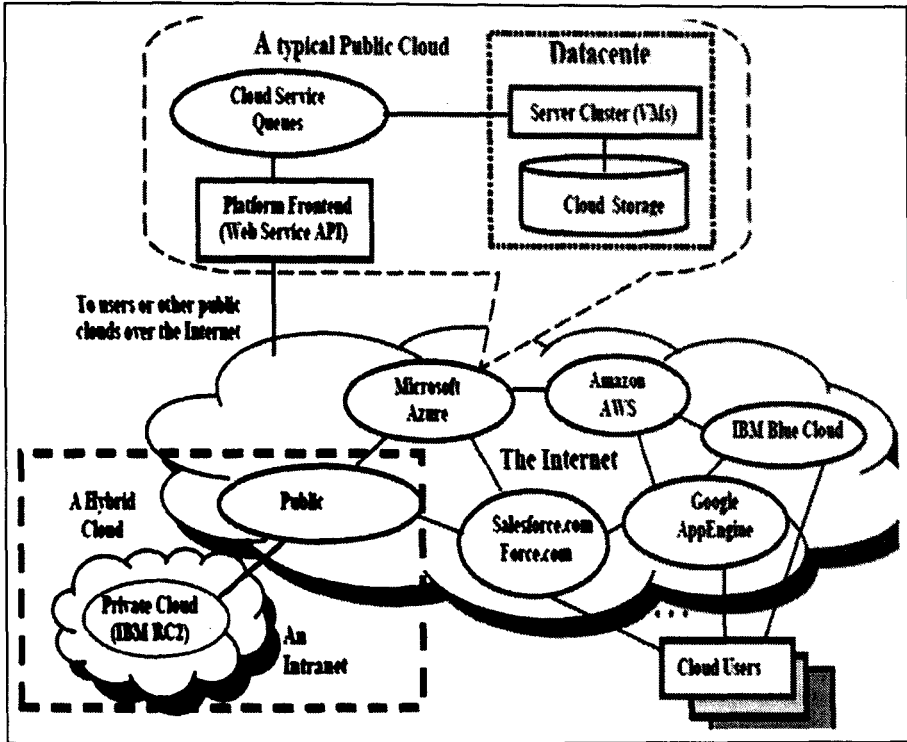


Fig. 12.3: Network Architecture of Hybrid Cloud

Role of Network in Cloud Security

Security is the top fear of IT managers when considering any cloud network service. Most IT managers are concerned about controlling access of their sensitive data . From a networking standpoint, each service model requires the cloud provider to expose more or less of the network and provide more or fewer networking capabilities to cloud users. Conversely, each service model requires cloud users to understand and design more or less of the network to which they are exposed. The network is most exposed in the IaaS model and least in the SaaS model. The essential technological difference between the deployment models is

derived from the networking relationship between the cloud user and the cloud provider. In a private cloud, the user and provider are within the same trusted network boundary. In a public cloud, they are on different networks. In a hybrid cloud, a secured connection may exist between the user's and provider's networks, or the user's network may extend into the provider's cloud (or the reverse). In a community cloud, the structure depends on the charter and architecture of the organizations operating the cloud. So network plays an important role in cloud security.

Role of Network in Cloud Application Performance

In cloud computing network performance is similar to application performance. Any delay in accessing any cloud application is due to the glitches in the cloud provider datacenter network.

"No network performance guarantees \Rightarrow no application predictability."

Wide area network connections between data centers and users can play a critical role in application performance - latency, packet loss and bandwidth all can influence the way users perceive application performance. And while some applications may perform adequately over a public Cloud infrastructure, many enterprise services won't. So the game changer becomes the private Cloud, and specifically Cloud-tuning capabilities.

Role of Network in Cloud Reliability

Networking as mentioned plays an important role in cloud security, application performance and in resiliency. When user gets services with best performance and with security, trust is build up between cloud user and cloud provider that leads to the reliable cloud infrastructure. So the role of network is very important in cloud reliability.

Issued with Networking in Cloud Security and Performance

First, public cloud access networking is most often through the Internet , though some cloud providers may be able to support virtual private networks (VPNs) for large customers. Accessing public cloud services will always create a tension between security and performance. Not all cloud computing providers will support encrypted tunnels, so your information may be sent in the open on the Internet. Where encryption is available, using it will certainly increase delay and may impact performance.

Our third network application in cloud computing is intra-cloud networking for public cloud services. Public cloud computing networks are internal and thus not visible to the user, so when you secure public cloud computing services, it is very important to understand how your provider interconnects its cloud

computing elements. The key issue to look for is the difference in network quality of service across the geography of the resource pool. If your cloud provider allows you to geographically narrow the range of resources that can be assigned to your application, then the performance variation across that narrower range should be examined. You'll want to include the intra-cloud network performance of public cloud providers in your cloud computing SLA.

Solutions to the Issues with Networking in Cloud Security Performance

Since several cloud providers don't use the encrypted tunnel for sending and receiving information over the internet, firstly encryption must be used at both cloud users and providers side. But as we know that encryption will bring delay to the communication of information application performance decreases thus to solve this we must reduce the number of transit hops that would not bring delay and would not affect the security. The Internet is a complex federation of interconnected providers, and reaching a given cloud computing service may involve transiting several provider networks. You will need to determine how your cloud provider choices are connected to other ISPs, particularly those you use regularly. The best cloud/ISP combination in terms of delay will almost always be one with the smallest number of hops. IT managers using or planning to use enterprise applications in a Cloud environment, which means just about every IT manager, need to have a heart-to-heart talk with their WAN service provider. Because network performance matters; architecture matters; adaptability matters and equally important, network design transparency matters. The problem of network security can be solved with the help of creating a VPN (Virtual Private Network) between users and cloud providers. A VPN is a network that uses a public communication infrastructure, such as the Internet, to provide remote offices or individual users with secure access to their organization's network. A VPN ensures privacy through security procedures and tunneling protocols such as the Layer Two Tunneling Protocol (L2TP). Data is encrypted at the sending end and decrypted at the receiving end. The data is encrypted by using various encryption algorithm such as MD5. These algorithms distort the actual message and convert it to something non-understandable so that data isn't intercepted in transit. To increase the security in cloud, intrusion prevention system, intrusion detection system are also installed. Sometimes they are the part of the firewall.

Conclusion

This research paper focused over several aspects of networking in cloud computing. The paper emphasized on the security, performance and reliability aspects of networking in cloud computing with network architecture of different deployment models. There are several issues related to security, performance, reliability that has been discussed in the paper that affects cloud computing,

possible solutions that are present today with various proposed solutions are also discussed in the paper. There is a lot of scope in the field of networking in cloud computing that need to be viewed seriously , scope over this is present that can be addressed later by different cloud and networking related people.

Acknowledgements

This research paper is made possible through the help and support from everyone including: parents, teachers, family, friends, and in essence, all sentient beings. We sincerely thank to our parents, family, and friends, who provide the advice and financial support. The product of this research paper would not be possible without all of them.

References

1. www.ibm.com/virtualization/networking
2. <http://www.cisco.com/go/cloud>
3. http://www.cs.rice.edu/~eugeneng/papers/INFO_COM10-ec2.pdf
4. <http://www.cse.wustl.edu/~jain/cse571-11/ftp/virtual/#sec2.1>
5. http://www.computerweekly.com/news/1280095_810/Cloud-computing-How-to-avoid-a-network-bottleneck
6. <http://www.informationweek.com/whitepaper/Infrastructure/Network-Systems-Management/the-impact-of-cloud-computing-on-the-network-wp1338307977?articleID=191704931>
7. <http://la.trendmicro.com/media/wp/cloud-based-threat-defense-model-esg-whitepaper-en.pdf>
8. <http://www.f5.com/pdf/white-papers/virtual-data-center-security-wp.pdf>

Collaborative Recommender Systems via User-Item Subgroups

*Sharmistha Kakati
Sarat Kr. Chettri*

Abstract

People often face dilemma in deciding upon certain matters and in that situation they feel the need of feedback from some other sources. In such cases, often, people rely upon word of mouth from other people who are familiar with their item of interest or advice from other people. Recommender systems were introduced to serve exactly this purpose and with the rise in number of internet users, it has become an inevitable part. In this paper we have proposed and developed an automated recommender system which studies the preferences of users over certain consumable/non-consumable items and cluster them on the basis of some similarity criterion. The system recommends user with new items based on the choices made by other users with similar interests.

Keywords : *Collaborative Filtering, Recommender Systems, Cold Start Problem.*

Introduction

Recommender system, also known as recommendation system or recommendation engine, are a type of information filtering systems which attempt to predict a user's preference over certain items (which the user had not considered before) by studying the characteristics of the target item or the user's social behaviour. The range of topics where we might need recommendations may also vary from a wide range of topics like which restaurant we would eat, which book we would buy or which movie we would like to watch. The paper

has been organized as; Section 2 elaborates related concepts on previous works done on recommender systems. Section 3 starts with a brief idea about the cold start problem and provides a layout of the three issues defined and solved during the course of this work. Section 4 explains the experimentation process and also states some of the common measures used for evaluation. Conclusions are drawn in Section 5 with some directions of future work.

Background and Related Concepts

Recommender systems date back to the eighties and since then many improvements have been made to provide us with better service quality. A lot of work has been done in this field by many research scholars as the advent of the internet has made recommender systems a trending topic for research. In 1979, Rich [15] developed the Grundy System which helped in building individual user models which in turn aided in giving very relevant books to the users. After that, the tapestry system was developed by Goldberg, Nicholas, Oki and Terry in 1992 in which like-minded users had to be found manually. Lee, Kim and Rhee have done an extensive research on traditional recommendation systems that exist and have found solutions to many problems which can be found in such systems. They showed that performance of recommender systems can be improved by analysing web access patterns. A very important thing which was pointed out in this paper was that content based filtering is unsuitable for modern day web [1]. Again Xu, Bu, Chen and Cai take a complete different approach to the problem. They introduced the concept of dividing the users and items into multiple subgroups. However one disadvantage in this process was that it proved to be too costly[2]. Bobadilla, Ortega, Hernando and Bernal[4] propose an innovative similarity metric:HwSimilarity. This metric attains high-quality recommendations that are similar to those provided by the best existing metrics and can be processed by employing low-cost hardware circuits. This paper examines the key design concepts and recommendation-quality results of the metric. This metric compares both the numerical similarity and non-numerical similarity between users.

There are many useful reviews that give us a good understanding of the trend of recommender systems[8]. The basic concept of recommender systems have changed drastically since the beginning. Earlier recommender systems included only content based, demographic and collaborative filtering. On the contrary now a days they incorporate in social information also [9] like in social networking sites; Facebook, LinkedIn, Twitter etc.. Some of the earlier systems which used collaborative filtering are Grouplens [15][16] which is a news and article recommender system, Bellcore's Video Recommender [19] which is a movie recommender and Ringo[18] which recommends music and related music artists to users. For achieving rating prediction, Non-negative Matrix Factorisation algorithms have been used to devise a unified model by Gu, Zhou, and Ding [20].

Artificial intelligence has also been incorporated in recommender systems by many researchers. Ujjin and Bently [15] incorporated a Particle Swarm Optimization recommender system in which the PSO algorithm was applied to study a user's personal preferences and provide suggestions to that user based on his/her preferences. Another system which uses collaborative filtering is the book recommendation system of Amazon.com [17]. Juan F. Huete, J.M. Fernández-Luna, Luis M. de Campos and Miguel A. Rueda-Morales [21] challenged the typical user based rating prediction system where each neighbour rates the target item and each neighbour's contribution are combined to weigh a suggestion. They proposed a model which is independent of user ratings and is based on predictive probabilities.

In 2007, tag based CF systems were introduced by Nakamoto, Nakajimi, Miyazaki and Uemura in their paper. They proposed a tag-based contextual CF by depicting overlapping of tags. They constructed two models which combined tags into user similarity calculation stage and score prediction stage. But a drawback of the system was that it had the possibility of inefficiency if there were not enough tags. Ji, Yeon, Kim and Jo in 2007 used tags to find similar users and used it to form candidate tag sets for each user, upon which they used Naïve Bayes classifier to implement recommendation approach. Zheng and Li, in 2011, build a model that used tag and time information to predict a user's taste.

Proposed Methodology

The present work deals with the cold start problem and tries to simplify it to a certain extent. The cold start problem is most prevalent in recommender systems. Recommender systems form a specific type of information filtering (IF) technique that attempts to present information items (movies, music, books, news, images, web pages) that are likely of interest to the user. Typically, a recommender system compares the user's profile to some reference characteristics. These characteristics may be from the information item (the content-based approach) or the user's social environment (the collaborative filtering approach).

In the content-based approach, the system must be capable of matching the characteristics of an item against relevant features in the user's profile. In order to do this, it must first construct a sufficiently-detailed model of the user's tastes and preferences through preference elicitation. This may be done either explicitly (by querying the user) or implicitly (by observing the user's behaviour). In both cases, the cold start problem would imply that the user has to dedicate an amount of effort using the system in its 'dumb' state - contributing to the construction of their user profile - before the system can start providing any intelligent recommendations.

In the collaborative filtering approach, the recommender system would

identify users who share the same preferences (e.g. rating patterns) with the active user, and propose items which the like-minded users favoured (and the active user has not yet seen). Due to the cold start problem, this approach would fail to consider items which no-one in the community has rated previously.

Figure 13.1 illustrates our proposed recommendation model based on user and item similarities. First the data collected is pre-processed and then clustered according to user similarities and item similarities. Based on that similarity, new users who will be classified and finally the recommendation task happens where the new users will be recommended items based on that similarity criterion. The proposed framework for our method is as follows:

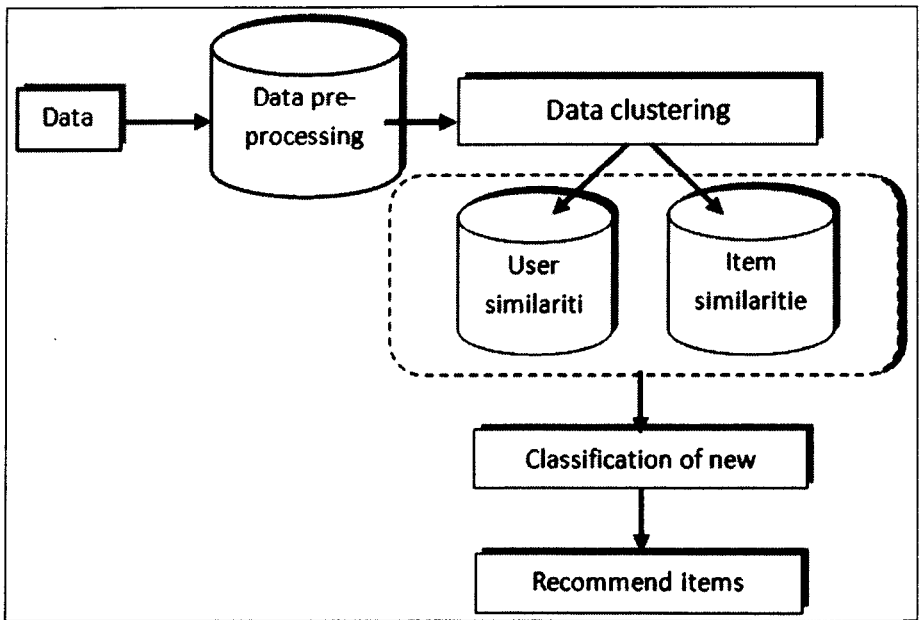


Fig. 13.1: Block diagram of the proposed system based on user

In order to deal with the cold start problem, we have identified three issues or scenarios which may arise when a person is dealing with a recommender system. The three issues are demonstrated below:

Issue 1

A customer has bought a book and has rated it. The age of the customer is known. How will a book be recommended to that customer? Suppose, a customer bought a book (book id is in the dataset) and rated it (Say 6). We compare the ratings which the customer has given to the book with the ratings

given by other users and find the range, under which the rating given by the new customer falls. Then we break the range into individual rating.

For every rating, we check how many users fall under each rating, and we find the most similar user with the customer who falls under the same age. Then we find what book that user has rated and we recommend that book to the new customer. Following are the parameters for the proposed algorithm:

X = New user

a_x = Age of X

B_i = Existing book id

B = New book with book id B_i

R_x = Rating given by X

R_i = Rating of books; i belongs to $\{1, 1+1, \dots, h\}$

U_i = Existing user id

Algorithm 1: Proposed algorithm for issue 1

Input: ISBN number, rating, user id, age

Output: System recommends the most similar book to the user according to the age of the user.

1. User X rates book B as R_x , $R = \{R_i, R_{i+1}, \dots, R_h\}$.
2. For book B , make bin range as $b_i - b_{i+2}, b_{i+3} - b_{i+5}, \dots, b_{h-1} - b_h$, where $0 \leq h$
3. Put existing users U_k in defined bin ranges according to ratings R_x for B .
4. Check for R_x in defined bin ranges.
5. Check for n users in defined bins where $n > 0$.
6. Find nearest neighbors of X among U_k with age a_x .
7. For every $R_{h-3} - R_h$ of B_i Y has rated, recommend B_i to X .

Issue 2

A customer has bought a book and has not rated it. The age of the customer is known. How will a book be recommended to that customer? Suppose, a customer bought a book and has not rated it yet. Find the relevant category of that book id and the age of the customer. Perform item based clustering based on age and category. Suppose, Category = "Mystery"

Age of user $r = 25$

For the particular age range make bins of the books having highest ratings.

Suggest the book with the highest rating. The parameters for our proposed algorithm are as follows:

X = New user

a_x = Age of X

B_i = Existing book id

C = category of B_i

Algorithm 2: Proposed algorithm for issue 2

Input: ISBN number, rating, user id, age, category

Output: System recommends the most similar book to the user according to the age of the user and the category to which the book belongs.

1. User X of age a_x buys a book B_i of category C .
2. Let n be the number of users who bought books in category C .
3. Choose from n , the number of users who have rated C category books within rating range R_{h-4} R_h .
4. Make separate bins b_{h-4} , b_{h-3} , b_{h-2} , b_h according to rating.
5. For all users falling in the defined bins, find nearest neighbors of X with age a_x .
6. Recommend books rated by those users to X

Issue 3

A book is newly arrived in the market and it has not been rated yet. How will the book be recommended to anyone? Suppose a book is newly arrived and has not been rated yet. We first find the category of the book. Then we find the users who have rated books falling under that same category and recommend the book to them.

B = New book

C = Category of the book

u_i = User preferring books of category C .

Algorithm 3: Proposed algorithm for issue 3

Input: ISBN number, category

Output: System recommends the most similar book to the user according to the category to which the book belongs.

1. Let C be a category of book B .
2. Let n be the number of users who bought books in category C .

3. Choose from n , the number of users who have rated C category books within rating range $R_{h-4} \dots R_h$.
4. Make separate bins $b_{h-4}, b_{h-3}, b_{h-2}, b_h$ according to rating.
5. For all users falling in the defined bins, say u_i .
6. Recommend book B to u_i .

Choosing the Similarity Metric

We have chosen the absolute distance as our similarity metric.

Absolute distance between two vectors gives us the vector nearest to the origin. If a and b are two vectors, then,

$$\text{Dist}(a, b) = \sqrt{\sum_{i=1}^n (a_i - b_i)^2}$$

Where $a = (a_1, a_2, \dots, a_n)$ and $b = (b_1, b_2, \dots, b_n)$.

Experimental Methodology

We have used the Book Crossing-dataset which is available online. This dataset was collected and compiled by Cai-Nicolas Ziegler in a 4-week crawl (August / September 2004) from the Book-Crossing community with kind permission from Ron Hornbaker, CTO of Humankind Systems. It contains 278,858 users (anonymized but with demographic information) providing 1,149,780 ratings (explicit / implicit) about 271,379 books. It is available on the site [23].

The Book-Crossing dataset comprises 3 tables.

BX-Users - This table contains the users. Note that user IDs ('User-ID') have been anonymized and mapped to integers. Demographic data is provided ('Location', 'Age') if available. Otherwise, these fields contain NULL-values.

BX-Books - Books are identified by their respective ISBN. Invalid ISBNs have already been removed from the dataset. Moreover, some content-based information is given ('Book-Title', 'Book-Author', 'Year-Of-Publication', 'Publisher'), obtained from Amazon Web Services. Note that in case of several authors, only the first is provided. URLs linking to cover images are also given, appearing in three different flavours ('Image-URL-S', 'Image-URL-M', 'Image-URL-L'), i.e., small, medium, large. These URLs point to the Amazon web site.

BX-Book-Ratings-this table contains the book rating information. Two types of ratings can be seen- 'Book Ratings' which are explicit and expressed on a scale from 1-10, and implicit expressed by 0.

The collected data were analysed to detect any anomalies that may be

present. To implement our proposed method, we require four attributes from among all the three tables. The required attributes are book id, rating, user id and age. We extract the required attributes from the downloaded datasets and prepare our own table consisting of the four attributes book id (ISBN), book rating, user id and age. The places containing null values were omitted.

To test our method, we make a test dataset of 200 data items from the original dataset. The dataset consists of the attributes book id and age of users. We make a separate dataset where we take the book ids of the first two hundred items.

Proposed Evaluation Methodology

In order to measure the accuracy of a recommender system, we need to find the following parameters:

True Positive: System recommended a book and the user likes it.

True Negative: System recommended a book and the user did not like it.

False Positive: System recommended a book and the user did not like it.

False Negative: System did not recommend the book but the user liked it.

There are different accuracy measures to calculate the accuracy of a system. A few of them which can be used to measure the accuracy of a recommender system are listed below:

TP Rate : The True Positive Rate (TPR) are those examples which truly have class x and were also classified as class x during experimentation. In the confusion matrix, this is the diagonal element divided by the sum over the relevant row.

$$\text{TP Rate} = \frac{TP}{TP + FN} \quad (4.1)$$

Where TP is the True Positive and FN is False Negative.

FP Rate: The False Positive Rate (FPR), are those examples which belong to a class y but are referred to as class x. In the confusion matrix, this is the column sum of class x minus the diagonal element, divided by the rows sums of all other classes.

$$\text{FP Rate} = \frac{FP}{FP + TN} \quad (4.2)$$

Where, FP is the False Positive and TN is True Negative.

Precision : Precision rate gives the values which actually belongs to class x and also classified to be from class x. In the confusion matrix, this is the diagonal element divided by the sum over the relevant column.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4.3)$$

Where TP is true Positive and FP is False Positive

Recall: Recall value, also known as sensitivity, is the fraction of relevant instances that are retrieved in a search.

F-Measure : The F-Measure is a combined measure for precision and recall.

$$\text{F-measure} = \frac{2(\text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})} \quad (4.4)$$

Conclusion

Recommender systems have become indispensable. People use them to find books, music, news, smart phones and vacation trips. Nearly every product, service, or type of information has recommenders to help people select from among the huge number of alternatives the few they would most appreciate. Customers purchasing products such as financial services, computers, or digital cameras and so on need both information and intelligent interaction mechanisms that support the selection of appropriate solutions. So, an explicit representation of product, marketing, and sales knowledge is necessary which is implemented explicitly in the recommender systems. Because of its vast increasing demand, recommender systems have become a promising area of advanced research.

Presently in our research work, we have identified three issues and have successfully implemented all three of them. Out of the three issues, issue 2 and issue 3 addresses the cold start problem. As for future work, we need to evaluate our methods and compare with the existing state-of-the-art methods in the same domain. The use of association rule mining for better recommendation results can also be looked upon as future work.

References

1. C.H. Lee, Y.H. Rim and P.K. Rhee. Web personalization expert with combining collaborative filtering and association rule mining. *Expert Systems with Applications* 21, 2001.
2. Bin Xu, Jiajun Bu, Chun Chen and Deng Cai. An Exploration of Improving Collaborative Recommender Systems via User-Item Subgroups. WWW 2012 - Session: Recommender Systems April 16-20, 2012.
3. Taek-Hun Kim, Young-Suk Ryu, Seok-In Park and Sung-Bong Yang. An Improved Recommendation Algorithm in Collaborative Filtering. *E-Commerce and Web Technologies, Lecture Notes in Computer Science*, Volume 2455, pp 254-261, 2012.
4. Jesús Bobadilla, Fernando Ortega, Antonio Hernando and Jesús Bernal. Generalization of recommender systems: Collaborative filtering extended to

- groups of users and restricted to groups of items, *Expert Systems with Applications* 39, 172-186, 2012.
5. Gediminas Adomavicius and Alexander Tuzhilin. Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions, *IEEE Transactions On Knowledge And Data Engineering*, Vol. 17, No. 6, June, 2005.
 6. Weiyang Lin, Sergio A. Alvarez, Carolina Ruiz. Efficient Adaptive-Support Association Rule Mining for Recommender Systems, *Data Mining and Knowledge Discovery*, 6, 83-105, 2002.
 7. Joseph A. Calandrino, Ann Kilzer, Arvind Narayanan, Edward W. Felten, and Vitaly Shmatikov. Privacy Risks of Collaborative Filtering, *IEEE Symposium on Security and Privacy*, 2011.
 8. DeukHee Park, HyeaKyeong Kim, Li Young Choi and Jae Kyeong Kim. A literature review and classification of recommender systems research, *Expert Systems with Applications* 39, 10059-10072, 2012.
 9. Wang, J. Language models of collaborative filtering. In *Proceedings of the 5th Asia information retrieval symposium on information retrieval technology. AIRS '09* (pp. 218-229). Berlin, Heidelberg: Springer-Verlag, 2009.
 10. Wang, J., de Vries, A., & Reinders, M. A user-item relevance model for log-based collaborative filtering. In *proceedings of the 28th European conference on information retrieval. ECIR'06* (pp. 37-48). Berlin, Heidelberg: Springer-Verlag, 2006.
 11. Wang, J., de Vries, A. P., & Reinders, M. J. T. Unifying user-based and item-based collaborative filtering approaches by similarity fusion. In *Proceedings of the 19th annual international ACM SIGIR conference on research and development in information retrieval. SIGIR'06* (pp. 501-508), 2006.
 12. Wang, J., de Vries, A. P., & Reinders, M. J. T. Unified relevance models for rating prediction in collaborative filtering. *ACM Transactions on Information and System*, 26(3), 16:1-16:42, 2008.
 13. Wang, J., Robertson, S., de Vries, A. P., & Reinders, M. J. T. Probabilistic relevance ranking for collaborative filtering. *Information Retrieval*, 11(6), 477-497, 2008.
 14. Ujjin, S., & Bentley, P.J. Particle Swarm Optimization recommender system, in *Proceedings of the 2003 IEEE Swarm Intelligence Symposium-SIS'03*, pp. 124-131. doi:10.1109/SIS.2003.1202257, 2003.
 15. Konstan, J. A., Miller, B. N., Maltz, D., Herlocker, J. L., Gordon, L. R., & Riedl, J. GroupLens: Applying collaborative filtering to UseNet news. *Communications of the ACM*, 40(3), 77-87. <http://doi.acm.org/10.1145/245108.245126>, 1997.
 16. Resnick, Lakovou, Sushak, Bergstrom, & Riedl. GroupLens: An open architecture for collaborative filtering of netnews, in *proceedings of the 1994 ACM conference on Computer supported cooperative work*, 175-186,
 17. Greg Linden, Brent Smith, and Jeremy York, *Amazon.com Recommendations: Item-to-Item Collaborative Filtering*, Industry Report.
 18. Shardanand, U., & Maes, P. Social information filtering: Algorithms for automating

- 'Word of Mouth', in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Denver, Colorado, United States, pp. 210-217, 1995.
19. Hill, Stead, Rosenstein, & Furnas. Recommending and evaluating choices in a virtual community of use, in Proceedings of the SIGCHI Conference on Human Factors In Computing Systems, 194-201, 1995
 20. Gu, Q., Zhou, J., & Ding, C. H. Q. Collaborative filtering: Weighted nonnegative matrix factorization incorporating user and item graphs. In Proceedings of the SIAM conference on data mining. SDM 2010(pp. 199-210), 2012.
 21. Juan F. Huete, J.M. Fernández-Luna, Luis M. de Campos, Miguel A. Rueda-Morales. Using past-prediction accuracy in recommender systems. Information Sciences 199, 78-92, 2012.
 22. Nakamoto, R., Nakajima, S., Miyazaki, J. & Uemura, S. Tag-based contextual collaborative filtering. In Proceedings of 18th IEICE data engineering workshop, 2007.
 23. Ji, A. -T., Yeon, C., Kim, H. -N., & Jo, G. -S. Collaborative tagging in recommender systems. In Advances in artificial intelligence (AI2007)pp. 377-386, 2007.
 24. Nan Zheng, Qiudan Li. A recommender system based on tag and time information for social tagging systems. In Expert Systems with Applications 38 4575-4587, 2011.
 25. Dataset available on the following website: <http://www2.informatik.uni-freiburg.de/~chiegler/BX/>

Application of First Order Fuzzy Time Series In Enrollment Forecasting Technique

*Jutang P. Swer
B. Borah*

Abstract

A new method for forecasting enrollment using time-invariant fuzzy time series is proposed in this paper. Forecasting methods using first order fuzzy time series had been proposed earlier in which historical data are linguistic values. In this paper a new method for enrollment forecasting based on clustering and weights have been presented. Historical data which is the historical enrollment of the University of Alabama is applied here. The proposed method shows improvement over some previous existing models which also used fuzzy time series time-invariant. Results and accuracy are compared based on Average forecasting error rate and Mean square error. The proposed method shows higher average forecasting accuracy rates than the existing methods.

Keywords: *Forecasting, Fuzzy Time Series, Time-Invariant, Fuzzy Logical Relationship.*

Introduction

Forecasting is the process of making statements about events whose actual outcomes (typically) have not yet been observed. It is commonly used to predict weather conditions, earthquake, car accident rate, stock market and many more. Many forecasting techniques use past or historical data in the form of time series. A time series is simply a set of observations measured at successive points in

time or over successive periods of time. The measurements may be taken every hour, day, week, month, or year, or at any other regular (or irregular) interval. There are many well-known forecasting techniques however they cannot solve forecasting problems, in which the historical data are available in linguistic form. Hence, we need fuzzy time series to remove this disadvantage [1]. Song and Chissom [2], [3] introduced the concept of time-invariant and time-variant using fuzzy time series.

Fuzzy Time Series

Here the concept fuzzy time series is discussed. Fuzzy time series was proposed in order to deal with such situations in which the traditional time series model is no longer applicable. For example, in a forecasting problem, the historical data are not real numbers but are linguistic values. In this case, no models of conventional time series can be applied but a fuzzy time series model can be applied more properly. There are two possibilities in any application cases. One is that the historical data are in the forms of linguistic values. The other is that the historical data are real numbers. In the second case, the data should be fuzzified first. In our study, the historical data are real numbers. Therefore, in this application problem, there is one more step than when the data are linguistic value - the fuzzification of data. The main difference between the fuzzy time series and conventional time series is that the values of the former are fuzzy sets while the values of the latter are real numbers. Roughly speaking, a fuzzy set is a class with fuzzy boundaries.

Let U be the universe of discourse,

$$U = \{u_1, u_2, \dots, u_n\}. \quad (1)$$

A fuzzy set A of U is defined by

$$A = f_A(u_1)/u_1 + f_A(u_2)/u_2 + \dots + f_A(u_n)/u_n \quad (2)$$

where f_A is the membership function of A , $f_A: U \rightarrow [0, 1]$, and $f_A(u_i)$ indicates the grade of membership of u_i in A , where $f_A(u_i) \in [0, 1]$ and $1 \leq i \leq n$. The definitions of fuzzy time series are reviewed as follows.

Definition 1

Let $Y(t)$ ($t = 0, 1, 2, \dots$), a subset of R , be the universe of discourse on which fuzzy sets $f_i(t)$ ($i = 1, 2, \dots$) are defined and let $F(t)$ be a collection of $f_i(t)$ ($i = 1, 2, \dots$). Then, $F(t)$ is called a fuzzy time series on $Y(t)$ ($t = 0, 1, 2, \dots$).

From Definition 1, we can see that $F(t)$ can be regarded as a linguistic variable and $f_i(t)$ ($i = 1, 2, \dots$) can be viewed as possible linguistic values of $F(t)$, where $f_i(t)$ ($i = 1, 2, \dots$) are represented by fuzzy sets. We also can see that $F(t)$ is a function of time t , i.e., the values of $F(t)$ can be different at different

times due to the fact that the universe of discourse can be different at different times. According to [1], if $F(t)$ is caused by $F(t-1)$ only, then this relationship is represented by $F(t-1) \rightarrow F(t)$.

Definition 1.1.

Let $F(t)$ be a fuzzy time series. If for any time t ,

$$F(t) = F(t-1) \quad (3)$$

and $F(t)$ only has finite elements, then $F(t)$ is called a time-invariant fuzzy time series. This implies that in the case of time-invariant fuzzy time series, it is very easy and convenient to calculate the first-order model. In this case it is actually assumed that $F(t)$ has the same possible values at any t . Therefore, the model is independent of time t .

Otherwise, it is called a time-variant fuzzy time series where

$$F(t) = F(t-1), F(t-2), F(t-3), \dots, F(t-n) \quad (4)$$

This means that there is a possibility that a fuzzy time series may not have the same possible values at different times.

New Approach to Forecast Enrollment Using Fuzzy Time Series

From the *actual* historical data of enrollments of the University of Alabama Song and Chissom [2], [3] set up models, i.e. relationships among values of interests at different moments of time. Likewise, Chen [1] puts up method for construction of fuzzy sets A_i being values of the linguistic variable (actual) enrollments. In this paper a new method for forecasting enrollment is presented. According to the proposed method, a clustering algorithm to cluster the historical data into intervals of different lengths is applied. The proposed method then constructs first-order fuzzy logical relationship rules. Then, two different weights are used for forecasting. The weights used are the frequency of interval (i.e. the number of times an interval occur) and the frequency of fuzzy logical relationship (i.e. number of times a fuzzy logical relationship occurs). These two weights are combined together in order to increase the forecasting accuracy rate of prediction. The new method is as follows:

Step 1: Sort the data in ascending order (we counted the same value only once). Assume that there are n numerical data. Suppose we have ascending numerical data as follows:

$$d_{1,0} = d_{1,1} = \dots < d_{2,0} = d_{2,1} = \dots < \dots < d_{n-1,0} = d_{n-1,1} = \dots < \dots d_{n,0} = d_{n,1} = \dots$$

where $d_{i,0}, d_{i,1}, \dots$ denote the numerical data with the same value, $1 \leq i \leq n$

Then we calculate the Eps (i.e average difference) as follows:

$$Eps = \frac{\sum_{i=1}^{n-1} (d_{i+1,0} - d_{i,0})}{n-1} \quad (5)$$

where "Eps" denotes the average of the differences between any two adjacent data.

Step 2: Using Eps, we determine whether two adjacent data can be put into a cluster. Suppose we take x_i and x_j as two adjacent data.

If $|x_i - x_j| \leq \text{Eps}$, then we put x_j into the cluster where x_i belongs to otherwise we create a new cluster for x_j . We keep on comparing every two adjacent data until all the data are clustered.

Step 3: We now calculate α which determines whether a cluster should be decomposed or not and also the maximum number of data to be present in a cluster after decomposition as follows:

$$\alpha = \text{int}(n / k) + 1 \quad (6)$$

where k is the number of clusters obtained after Step 2.

Using α we can determine whether a cluster should be decomposed or not. Let m be the number of data in a big cluster C_i , $1 \leq i \leq n$. Suppose for cluster C_i , $m > \alpha$, then we have to decompose cluster C_i . We can decompose C_i into the following number of clusters using the following conditions:

- If $(m \% \alpha) = 0$, then we decompose C_i into (m/α) clusters.
- If $(m \% \alpha) > 0$, then we decompose C_i into $\text{int}(m / \alpha) + 1$ clusters.

Step 4: Let the new number of clusters obtained after Step 3 be k' . This will be the number of intervals used. We define the intervals by taking the smallest data in a cluster as lower-bound and the biggest data as upper-bound of the interval. If there is only a single data in a cluster then we temporarily treat it as the lower-bound and upper-bound of the interval created.

For Example, cluster C_i and its adjacent cluster C_{i+1} can be converted into intervals $u_i = [lb_i, ub_i]$, $u_{i+1} = [lb_{i+1}, ub_{i+1}]$, $1 \leq i \leq n$, where lb_i , lb_{i+1} and ub_i , ub_{i+1} are the lower-bounds and upper-bounds of interval u_i and u_{i+1} respectively. Suppose cluster $C_j = \{x_j\}$, $1 \leq j \leq n$, then we convert C_j into interval $u_j = [x_j, x_j]$.

Here, the intervals obtained are not contiguous in nature. To make the intervals contiguous we take the mid-point of ub_i of interval u_i and lb_{i+1} of interval u_{i+1} and use this mid-point for new intervals w_i and w_{i+1} as follows:

$$w_i = [lb_i, (ub_i + lb_{i+1})/2], \quad w_{i+1} = [(ub_i + lb_{i+1})/2, ub_{i+1}]$$

Step 5: We now define the universe of discourse U . Let $U = [D_{\min}, D_{\max}]$ where D_{\min} and D_{\max} are the minimum value and the maximum value of the historical data. Based on the proposed approach, we divide the universe of discourse U into different lengths of intervals w_1, w_2, \dots, w_n .

Step 6: We define each linguistic term A_i represented by a fuzzy set, where $1 \leq i \leq q$, shown as:

$$A_1 = 1/w_1 + 0.5/w_2 + 0/w_3 + \dots + 0/w_n$$

$$A_2 = 0.5/w_1 + 1/w_2 + 0.5/w_3 + \dots + 0/w_n$$

$$A_q = 0/w_1 + 0/w_2 + 0/w_3 + \dots + 0.5/w_{n-1} + 1/w_n$$

where A_1, A_2, \dots, A_n are linguistic terms to describe the values.

Step 7: Fuzzify the historical data into fuzzy sets respectively. If the historical data belongs to w_i where $1 \leq i \leq n$, then we fuzzify the historical data into fuzzy set A_i .

Step 8: Construct the fuzzy logical relationships based on the fuzzified historical data obtained in Step 7. If fuzzified data of years t and $(t+1)$ are A_j and A_k respectively, then construct the fuzzy logical relationship " $A_j \rightarrow A_k$ ", where A_j is the current state and A_k is the next state of the fuzzy logical relationship. According to the current states of the fuzzy logical relationships, divide the fuzzy logical relationships into fuzzy logical relationship groups, where the fuzzy logical relationship having the same current state are put into the same fuzzy logical relationship group. We take only one ($A_j \rightarrow A_k$) fuzzy logical relationship in a fuzzy logical relationship group (in case there is more than one fuzzy logical relationship).

Step 9: Calculate the forecasted output as follows:

- i) If fuzzified data of year t is A_j and there is only one fuzzy logical relationship in the fuzzy logical relationship group whose current state is A_j shown as

$$A_j \rightarrow A_k$$

Then the forecasted value of year $(t+1)$ is m_k , where m_k is the mid-point of the interval w_k and the maximum membership value of fuzzy set A_k occurs at the interval w_k .

- ii) If fuzzified data of year t is A_j and there are the following fuzzy logical relationships in the fuzzy logical relationship group whose current state is A_j shown as follows:

$$A_j \rightarrow A_{k1}(x_1, y_1)$$

$$A_j \rightarrow A_{k2}(x_2, y_2)$$

$$A_j \rightarrow A_{kp}(x_p, y_p)$$

Then the forecasted value of year (t+1) is calculated as follows:

$$[m_{k_1} (x_1+y_1) + m_{k_2} (x_2+y_2) + + m_{k_p} (x_p + y_p)] / (x_1+y_1 + (x_2+y_2)+....+(x_p+y_p)$$

where x_i denotes the frequency of interval w_i (i.e. the number of times interval w_i occur), y_i denotes the number of fuzzy logical relationships

“ $A_j ! A_{ki}$ ” that occur in the fuzzy logical relationship group, $m_{k_1}, m_{k_2}, ..., m_{k_p}$ are the mid-points of the intervals $w_{k_1}, w_{k_2}, ..., w_{k_p}$ respectively and the maximum membership values of fuzzy sets $A_{k_1}, A_{k_2}, ..., A_{k_p}$ occur at the intervals w_{k_1}, w_{k_2} and w_{k_p} respectively where $1 \leq i \leq p$.

Results

The data set tested is the Historical enrollments of the University of Alabama from 1971 to 1992 collected from [1]. The Average forecasting error rate (AFER) and Mean square error (MSE) to compare the forecasting results of different forecasting methods is used, where A_i denotes the actual enrollment and F_i denotes the forecasting enrollment of year i , respectively. The experimental results of Enrollment forecasting according to the proposed method using First Order Fuzzy Time Series is depicted in Table 14.1 and 14.2.

$$AFER = \frac{|A_i - F_i| / A_i}{n} \times 100\% \tag{7}$$

$$MSE = \frac{\sum_{i=1}^n (A_i - F_i)^2}{n} \tag{8}$$

Table 14.1: Actual Enrollment and Forecasted Enrollment according to the proposed method

Years	Actual Enrollment	Forecasted Enrollment according to the proposed algorithm	Error
1971	13055	-	-
1972	13563	13512	51
1973	13867	13998	131
1974	14696	14600	96
1975	15460	15461	1

1976	15311	15286	25
1977	15603	15547	56
1978	15861	16533	672
1979	16807	16533	274
1980	16919	17136	217
1981	16388	17136	748
1982	15433	15461	28
1983	15497	15286	211
1984	15145	15286	141
1985	15163	15547	384
1986	15984	15547	437
1987	16859	16533	326
1988	18150	17136	1014
1989	18970	18831	139
1990	19328	19243	85
1991	19337	19037	300
1992	18876	19037	161

Table 14.2. A comparison of the Enrolment forecasting results of different forecasting methods using First Order

	Song Chissom [2]	Chen [1]	Sah and Degtiarev [4]	Proposed Method
MSE	775687	321418	-	135549
AFER	4.38 %	3.23 %	2.42 %	1.57402%

Conclusion

In this paper, a new method for forecasting based on clustering and weights have been presented. Each value in a real-valued time series is fuzzified into a fuzzy value represented by a fuzzy set which form a fuzzy time series. Therefore, the real-valued time series is transformed into a fuzzy time series, and then the fuzzy time series to model the real-valued time series. The proposed method applies a clustering algorithm to cluster the historical data into intervals of different lengths. Then two different weights are combined together in order to increase the forecasting accuracy rate of prediction. The proposed method gives higher average forecasting accuracy rates than the existing methods.

References

1. S.M. Chen, "Forecasting Enrollments Based on Fuzzy Time Series", *Fuzzy Sets and Systems*, vol. 81, pp. 311-319, 1996.
2. Q. Song and B.S. Chissom, "Forecasting enrollments with fuzzy time series – part 1", *Fuzzy Sets and Systems*, vol. 54, pp. 1-9, 1993.
3. Q. Song and B.S. Chissom, "Forecasting enrollments with fuzzy time series – part 2", *Fuzzy Sets and Systems*, vol. 62, pp. 1-8, 1994.
4. S. Melike, K. Y. Degtiarev, *Forecasting Enrollment Model Based on First-Order Fuzzy Time Series*, *Proceedings of World Academy of Science, Engineering and Technology*, 2005.

Analysis of Distributed Data with Preserved Privacy

Sarat Kumar Chettri

Abstract

Data mining relates to extracting knowledge from large amount of data. With the advent of latest data mining techniques, preserving the privacy of individual's data became a persistent issue. Every day tremendous amount of data is being generated electronically with increasing concern of data privacy. The data may be distributed among various parties and certain privacy and legal issues may prevent them to share their data for analysis. The main challenge is how parties collaboratively perform data analysis without breaching the privacy of their local data. In this paper we have used the data-oriented micro-aggregation method *MDAV2k* with semi-trusted third party protocol for microdata protection. The datasets under consideration are vertically partitioned among different sites. The performance comparison of the method has been done with reference to local and distributed datasets in terms of data utility and risk of data disclosure. Experimental results prove our claim and shows that the performance of *MDAV2k* method proves to be effective in dealing with vertically partitioned numerical dataset.

Keywords: *Microaggregation, vertically partitioned data, data utility, data disclosure risk, semi-trusted third party protocol.*

Introduction

Data mining techniques applied on the electronic data has been looked upon as a threat to the privacy of individuals as the individual's privacy may be breached if the databases related to them get disseminated. Much information related to

individuals contained in the databases may be sensitive in nature and if got revealed may lead to various legal and ethical issues. As far as privacy of the individuals is concerned, personal data are supposed to be kept private and there exist several kinds of data disclosure risks [1]. But this should not curtail the knowledge discovery process. Thus the main challenge is to apply various data mining techniques for data proliferation with preserved privacy. Advances in cryptography and information hiding meets the challenge but efficiency is one expect to be looked on as well advancement of data mining techniques has created recent interests in this area. A popular and efficient approach for preserving individual privacy and resisting linked attack is to anonymize data using k -anonymity. Samarati [15] and Sweeney [16] proposed a definition of privacy called k -anonymity. In this context a new method called micro-aggregation has been proposed in the literature. Micro-aggregation is a family of Statistical Disclosure Control (SDC) [9] method which naturally satisfies k -anonymity. For being the method perturbative in nature, the challenging task is to tune the modification of data in such a way that both privacy risk and information loss caused due to data modification are kept below certain acceptable limits. Various micro-aggregation techniques (fixed-size and variable-sized) exist in the literature [2, 3, 5, 6, 11, 13].

MDAV2k [3] is an efficient multivariate variable-sized micro-aggregation technique for microdata protection i.e. protection of data related to individual or organization. However, the method works only with the numerical [3] and mixed data [4] which are located centrally. In real life the data may be distributed among various parties and integrating the local datasets may give a globalized dataset. Certain issues may prevent the individual parties to share their local data to the other parties. Generally data may be distributed randomly, but such random distribution is rare. In this paper we have taken vertically partitioned data. If a dataset has been distributed among multiple sites S_1, S_2, \dots, S_n , the major problem is how each party collaboratively performs data analysis without revealing their data to each other thus maintaining the privacy of their privately owned datasets. To overcome the issue, we have extended the *MDAV2k* method for protecting microdata in this paper. In vertical partitioning, individual records of a dataset are spread out across different sites in such a way that each site may have different attributes for the same set of records. Related work exists in the literature [10, 12, 17, 18, 20], but to the best of our knowledge data-oriented micro-aggregation method has not been used for data privacy preservation, in the case that the data is vertically partitioned across different sites/parties.

The rest of the paper is organized as follows. Section 2 gives some related concepts and problem definition in conjunction with micro-aggregation of vertically partitioned data. In Section 3, the *MDAV2k* with semi-trusted third party protocol algorithm is presented. In Section 4, measures to evaluate proposed method are given. In Section 5, experimental data and results are presented and

the effectiveness of the proposed method is assessed. Finally, in Section 6 conclusions with direction for future works are drawn.

Micro-aggregation Techniques

Related Concepts

The main concept of micro-aggregation is to modify the dataset before its dissemination by making data at least k anonymous. It naturally satisfies k -anonymity without generalization and suppression as in [16]. The following concepts are required to understand the k -anonymity and micro-aggregation.

Definition 1. (Quasi-identifier)

A quasi-identifier (QI) is a set of attributes in a dataset D which can be used to identify a specific individual in D with the help of some external information. For e.g. job type, age, gender etc.

Definition 2. (k -anonymity)

A dataset D is said to be k -anonymous for $k > 1$ if, for any combination of values of quasi-identifiers (QI) there are at least k records indistinguishable from at least $(k-1)$ other records.

The main goal of micro-aggregation method is to produce a micro-aggregated dataset D' from an original dataset D in such a way that the data is protected from being disclosed to an intruder and at same time the user analyses on D and D' yields same or similar results.

Problem Definition

Analysis of vertically partitioned data is a complex task because different parties may have different data which may contribute additional information about the entities. The complexity significantly increases when the data analysis has to be done with preserved privacy abiding the legal restrictions of disclosure of any individual data. In our example a dataset D as shown in table I is vertically partitioned between two different sites S_1 and S_2 . Each site S_1 and S_2 contains a vertically partitioned dataset D_1 and D_2 respectively of the original dataset D . We denote the set of attributes owned by S_i , $i=1\dots n$ as $A_i = \{a_{i1}, a_{i2}, \dots, a_{im}\}$. D_1 consists of attributes *ID*, *race*, *date of birth* and *gender*. D_2 consists of *ID*, *zip*, *marital status* and *disease*. Tables II(a) and II(b) shows a vertical partitioning of the sample dataset D . The major problem is how each party P_1 and P_2 in sites S_1 and S_2 collaboratively performs data analysis without revealing their data to each other thus maintaining the privacy of their owned datasets.

Table 15.1. Original Dataset D

ID	Race	Date of Birth	Gender	ZIP	Marital Status	Disease
SS101	Asian	12/2/84	M	17901	Single	Diabetic
SS102	Asian	22/9/75	F	17902	Divorced	Flu
SS103	Asian	18/6/79	F	17903	Married	Diabetic
SS104	Asian	17/6/76	M	17904	Married	Malaria
SS105	Asian	12/4/84	M	17905	Married	Hypertension
SS106	Black	19/11/87	M	17906	Single	Asthma
SS107	Black	15/8/80	F	17907	Single	Obesity
SS108	White	22/7/79	F	17908	Single	Flu
SS109	White	19/5/81	F	17909	Widow	Malaria

Table 15.2. Vertical Partitioning of Dataset D between Two Different Sites

(a). Site S1

ID	ZIP	Marital Status	Disease
SS101	17901	Single	Diabetic
SS102	17902	Divorced	Flu
SS103	17903	Married	Diabetic
SS104	17904	Married	Malaria
SS105	17905	Married	Hypertension
SS106	17906	Single	Asthma
SS107	17907	Single	Obesity
SS108	17908	Single	Flu
SS109	17909	Widow	Malaria

(b). Site S2

ID	Race	Date of Birth	Gender
SS101	Asian	12/2/84	M
SS102	Asian	22/9/75	F
SS103	Asian	18/6/79	F
SS104	Asian	17/6/76	M
SS105	Asian	12/4/84	M
SS106	Black	19/11/87	M
SS107	Black	15/8/80	F
SS108	White	22/7/79	F
SS109	White	19/5/81	F

A common solution to this Secure Multi-party Computation problem [19] is to have a trusted third party who is trusted unconditionally. The solution is efficient with low communication cost but in such a scenario if the third party is dishonest then the privacy of the data provided by the communicating parties would be compromised. In this paper we have made an attempt to tackle the issue, where analysis of vertically partitioned data with privacy preservation is done even if the trusted third party becomes dishonest. The main idea followed here is that parties P_1 and P_2 communicates their partial micro-aggregated datasets and encrypted record identifiers to the semi-trusted third party P to perform some related computations and produce a modified dataset D' . Public-key cryptography is used to maintain the confidentiality.

MDAV2k Method Using Semi-Trusted Third Party Protocol

MDAV2k as seen in [3] is a variable-sized micro-aggregation method which partitions the dataset into groups of variable sizes depending on the distribution of the original records. The main issue tackled in the method is the determination of the gain factor \tilde{a} , which is computed dynamically to conservatively expand the group. The value for \tilde{a} checked and in case the value is greater than 1, then it is recomputed as $1+1/(5+\tilde{a})$ so as to restrict its value to a maximum of 1.67. The restriction is made to conservatively expand the group reducing information loss caused due to replacing the original values of the records by the centroid of the group to which it belongs. The complexity of the method has been proved to be $O(n^2)$. In this paper we have extended the *MDAV2k* method with semi-trusted third party protocol to protect vertically partitioned data among two different parties. The algorithm is given as under

Algorithm 1: *MDAV2k* Using Semi-trusted Third Party Protocol

1. Dataset D is partitioned vertically as D_1 and D_2 among two parties P_1 and P_2 . Each party P_1 and P_2 uses *MDAV2k* method to microaggregate their individual dataset locally. Based on this locally microaggregated data, a set S is produced containing Ids of the records partitioned into different groups. Each party then transmits the microaggregated dataset D'_1 and D'_2 along with encrypted form of S using public-key cryptography to the learning party P (semi-trusted third party). P decrypts S , then forms the modified dataset D' by joining the microaggregated datasets as $D' = D'_1 \bowtie D'_2$, if $S^i_a = S^j_b$ and $\{i \neq j\} = 1 \dots n$.
-

The basic concept followed in algorithm 1 is to modify the local datasets by the individual parties locally as shown in step 2. The modification is done using the *MDAV2k* method [3]. The parties P_1 and P_2 also produce a set S which consists of the unmodified Ids of the records. The set is encrypted using public-key cryptography to maintain the confidentiality. The set enables the semi-trusted third party P as shown in step 5 to integrate the different modified attributes of the same set of records, thus producing the globalized micro-aggregated dataset D^* . Algorithm 2 with its graphical representation (Fig. 1) shows the various steps performed to modify the dataset.

In the *MDAV2k* method, the similarity between any two records is found using the standard Euclidean distance and the centroid is computed as

$$Z_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} \quad (1)$$

Where Z_i is the centroid of group g_i , n_i is the number of numerical attributes in cluster g_i and x_{ij} is the j -th attribute of i -th group.

Algorithm 2: Maximum Distance to Average Vector (*MDAV2k* Method)

1. Locate the most distant record x_r from the global centroid \bar{x} of the dataset D . Find $2k$ nearest records of x_r , say $(y_1, y_2, \dots, y_{2k})$ and form a group g_i around x_r with its $(k-1)$ nearest records. Compute centroid \bar{x}_i of the formed grouped g_i and compute distances d_i from x_i to x_r and d_j from x_i to y_j where $j = k$. From y_j locate $(k-1)$ nearest records and compute a centroid \bar{z} of the temporary formed grouped around y_j . Find distance d_j from x_r to y_j and then compute the gain factor $\tilde{a} = d_j/d_i$. Set $\gamma = 1 + 1/(5 + \gamma)$ if $\gamma > 1$. Record y_j is inserted into group g_i if $(d_j < \gamma d_i)$. Check if $|g_i| < 2k$ then go to step 3 taking next $(k+1)$ nearest record of x_r . Repeat through step 1 if there remains at least $3k$ records to form any group. If there remains at least $2k$ records not belonging to any group then repeat the steps 1 and 2. 10. Form a new group with the remaining records.

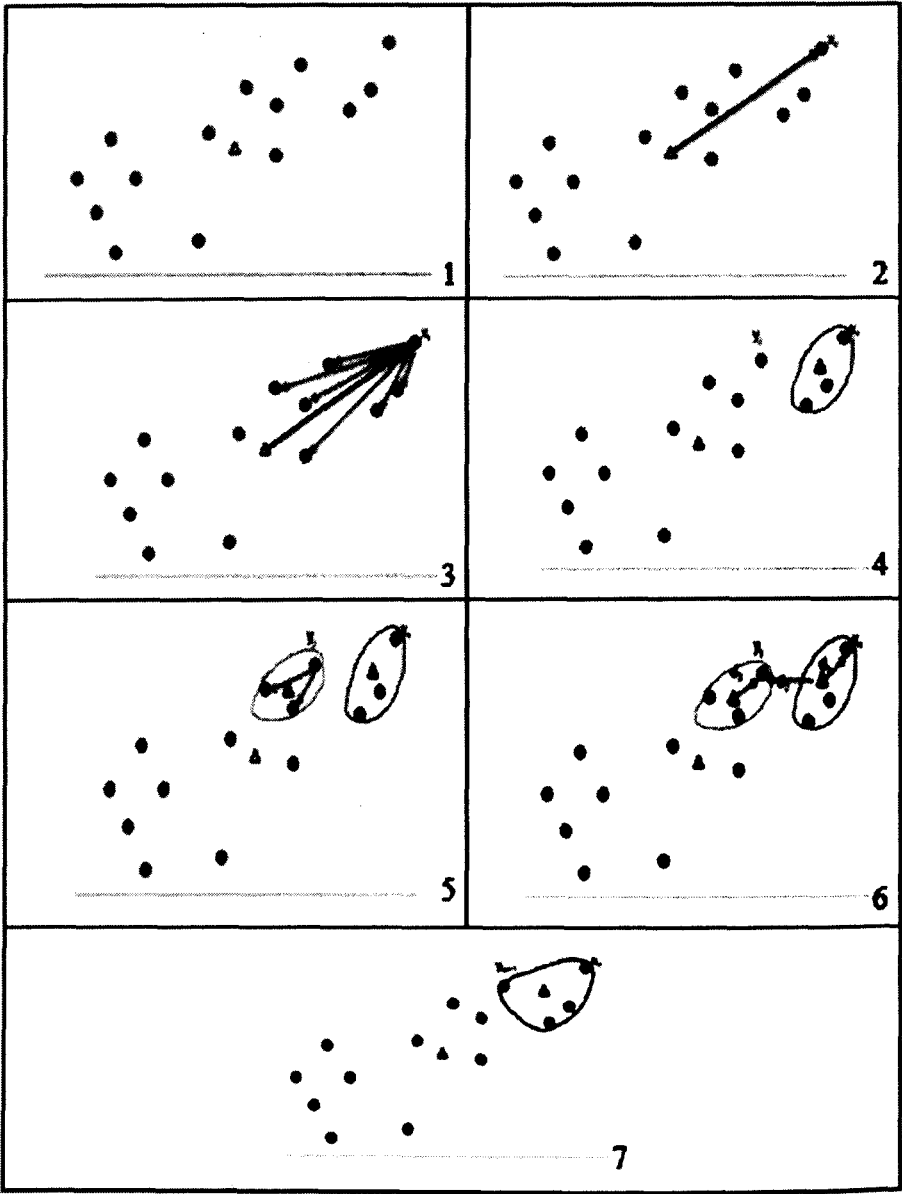


Fig. 15.1: Stepwise graphical representation of $MDAV2k$ method where dots represents record, arrow represent distance measurement and triangle represents centroid ($k = 3$).

Parameters for Proposed Method Evaluation

Information Loss Measure

The micro-aggregation method first partitions the dataset into groups and then micro-aggregates the records within the group by replacing the records with the centroid of the group. Thus it is said to be optimal if it partitions the dataset into k -partition with maximum within-groups homogeneity as can be seen in [8]. Higher within group homogeneity implies lower information loss, which means more data utility. As can be seen in [7] the sum of square error (SSE) parameter is commonly used to measure homogeneity in any formed group. The goal of micro-aggregation is to minimize the SSE measure, which is defined as

$$SSE = \sum_{i=1}^n \sum_{x_{ij} \in g_i} (x_{ij} - \bar{x}_i)' (x_{ij} - \bar{x}_i) \quad (2)$$

Where n is the total number of groups, g_i is the i -th group and \bar{x}_i is the centroid of g_i .

The total sum of square SST is the sum of square error within the entire dataset. SST represents the summed up distance of each record x_{ij} to the centroid \bar{x} of dataset D , which is defined as

$$SST = \sum_{i=1}^n \sum_{x_{ij} \in g_i} (x_{ij} - \bar{x})' (x_{ij} - \bar{x}) \quad (3)$$

The value SST for a dataset remains fixed while SSE changes with dataset partitioning into different groups. These two parameters are used to find the amount of loss of information in the dataset caused by micro-aggregating the records in individual formed groups. Information loss IL measure is defined as

$$IL = IL_1 + IL_2 \quad (4)$$

Where $IL_1 = \frac{SSE_1}{SST_1} \cdot 100$ and $IL_2 = \frac{SSE_2}{SST_2} \cdot 100$, SSE_1 and SST_1 are measured by P_1 , while SSE_2 and SST_2 are measured by P_2 . The parameter IL shown in equation 4 gives the total information loss caused due to data modification of dataset D .

Data Disclosure Risk

Once the dataset is modified, the security of anonymized datasets then needs to be accessed by measuring its data disclosure risk. Here in this paper the Distance Linkage Disclosure Risk (DLD) model [14] is used to analyze the probability of inferring the original record from the anonymized dataset. It can

be defined as for any anonymized record R' in an anonymized dataset D' if we compute a distance to other records in the original dataset D , we may get a nearest record R_1 and a second nearest record R_2 . If R_1 or R_2 is the original record R in the dataset D , then the record R is called a linked record. Let no_linked_record be the number of linked records in D' , $total_no_record$ be the total number of records in D' and then DLD is computed as

$$DLD = \frac{no_linked_record}{total_no_record} .100 \quad (5)$$

Experimental Data and Results

We have implemented the proposed method in C under Linux environment on a machine with 2.13 GHz Intel i3 processor and 3 GB RAM. Experiments are performed on the three benchmark datasets used during the “CASC” project [21]. The “Tarragona” dataset consists of 834 records with 13 numerical attributes. The “Census” dataset consists of 1,080 records with 13 numerical attributes and the “EIA” dataset contains 4,092 records with 11 numerical attributes. Data are standardized by having their mean μ subtracted from individual attribute values and then divided by the standard deviation σ of the population. This is done to ensure that all attributes have equal weights when computing distances. The algorithm is analyzed based on cluster homogeneity SSE as in equation 2, information loss measures IL as in equation 4 and DLD measures as in equation 5 with different values of k and the same parameters are used to compare the method against the $MDAV2k$ method when worked with the centralized data. The standard datasets are vertically partitioned among two different parties with a semi-trusted third party who integrates the partially microaggregated datasets to produce D' .

Fig. 15.1 compares the performance of the $MDAV2k$ method with the vertically partitioned data and centralized data. If the results are observed, it is clear that the performance of $MDAV2k$ with semi-trusted third party protocol method remains on par with the $MDAV2k$ method working with centralized data. The figure gives the comparison based on IL and DLD of the “Tarragona” and “Census” datasets. Fig. 15.2 gives the comparison based on IL and DLD of the “EIA” d

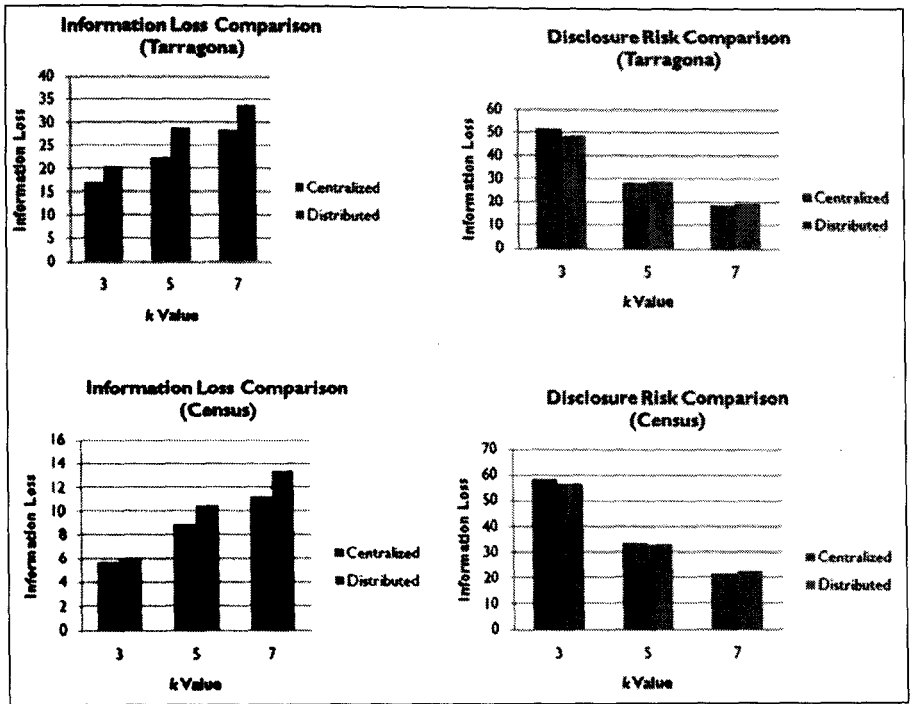


Fig. 15.2: Information Loss and Disclosure Risk Comparison for Different k Values for Tarragona and Census Datasets.

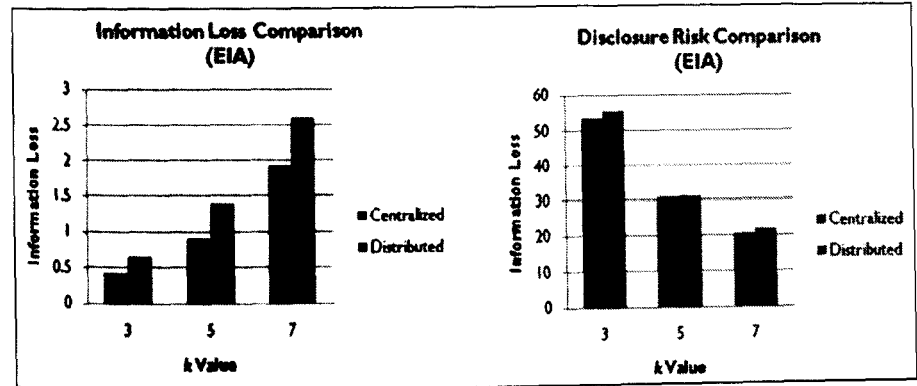


Fig. 15.3: Information Loss and Disclosure Risk Comparison for Different k Values for EIA Dataset.

Conclusion

In this paper we have extended the multivariate variable-sized micro-aggregation method *MDAV2k* to preserve the privacy of vertically partitioned datasets among two different parties with a semi-trusted third party. Experimental analyses with a publicly available standard datasets were performed to prove the efficiency of the method. Comparison has been done in terms of data utility and data disclosure risk when data is located centrally and distributed vertically. The experimental results prove that even a dataset is vertically distributed among multiple sites, secure analysis of dataset can still be performed with accuracy and preserved privacy without significant increase of the computational complexity. As a future work, the method may be extended towards privacy preservation of horizontally partitioned data. The method may also be extended towards preserving the privacy of time-series data.

References

1. Aggarwal, C.C., Pei, J., and Zhang, B., On privacy preservation against adversarial data mining. In: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '06, New York, USA, ACM Press; 2006. p. 510–516.
2. Chang, C.C., Li, Y.C., Huang and W. H. TFRP, An efficient micro-aggregation algorithm for statistical disclosure control. Journal of Systems and Software 80(11):1866–1878.
3. Chettri, S.K. and Borah, B., MDAV2K: A variable-size micro-aggregation technique for privacy preservation. In: International Conference on Information Technology Convergence and Services. Bangalore; 2012. p. 105–118.
4. Chettri, S.K. and Borah, B., An efficient micro-aggregation method for protecting mixed data. In: Computer Networks & Communications (NetCom). Springer, New York; 2013. p. 551–561.
5. A. Solanas and A. Mart'ínez-Balleste, V-MDAV: A multivariate micro-aggregation with variable group size, Seventh COMPSTAT Symposium of the IASC, Rome, 2006.
6. Domingo-Ferrer, J., Privacy in statistical databases: k-anonymity through micro-aggregation. IEEE Granular Computing, Atlanta; 2006. p. 747–777.
7. Domingo-Ferrer, J., Martinez-Balleste, A., Mateo-sanz, J. M. and Sebé, F., Efficient multivariate data-oriented micro-aggregation. The VLDB Journal, 2006: 355–369.
8. Domingo-Ferrer, J. and Mateo-Sanz, J.M., Practical data-oriented micro-aggregation for statistical disclosure control. IEEE Transactions on Knowledge and Data Engineering, 2002; 14(1): 189–201.
9. Fayyumi, E., A survey on statistical disclosure control and micro-aggregation techniques for secure statistical databases. Software: Practice and Experience, 2010; 40:1161–1188.
10. Karr, A.F., Lin, X., Sanil, A.P. and Reiter, J.P., Privacy-preserving analysis of

- vertically partitioned data using secure matrix products. *Journal of Official Statistics*, 2009; 25(1).
11. Lin, J. and Wen, T., Density-based micro-aggregation for statistical disclosure control. *Expert Systems with Applications*, 2010; 37(4):3256–3263.
 12. Mangasarian, O.L., Privacy-preserving linear programming. *Optimization Letters*, 2011; 5(1):165–172.
 13. Nin, J., On the disclosure risk of multivariate micro-aggregation. *Data and Knowledge Engineering*, 2008.
 14. Pagliuca, D., Some results of individual ranking method on the system of enterprise accounts annual survey. Esprit SDC Project, Deliverable MI-3/ D, 1999.
 15. Samarati, P., Protecting respondents identities in microdata release. *IEEE Transactions on Knowledge and Data Engineering*, 2001;1–29.
 16. Sweeney, L. K., Anonymity: A Model for Protecting Privacy. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 2002; 10(5):1–14.
 17. Vaidya, J. and Clifton, C., Privacy preserving association rule mining in vertically partitioned data. In: *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, 2002; 639–644.
 18. Vaidya, J., Clifton, C., Kantarcioglu, M. and Patterson, A.S., Privacy-preserving decision trees over vertically partitioned data. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 2008; 2(3).
 19. Yao, A.C.C., Protocols for secure computations. In: *FOCS*, vol. 82; 1982. p. 160–164.
 20. Zhu, X., Liu, M. and Xie, M., Privacy-preserving affinity propagation clustering over vertically partitioned data. In: *Fourth International Conference on Intelligent Networking and Collaborative Systems (INCoS)*, 2012, IEEE Computer Society, p. 311–317.
 21. Hundepool, A., de Wetering, V., Ramaswamy, R., Franconi, L., Capobianchi, A., DeWolf, P., Domingo-Ferrer, J., Torra, V., Brand, R. and Giessing, S., μ -ARGUS version 3.2 Software and User's Manual, Voorburg NL: Statistics Netherlands, 2003, <http://neon.vb.cbs.nl/casc>.

Sentiment Analysis of Twitter Data Base on Hashtag

*Stevenson Mawa
N Donald Jefferson Thabah*

Abstract

Twitter is one of the largest social networking website that receives tweets in millions every day. These tweet messages are short and are generated constantly. These tweet data can be used to analyze and visualize for business or industrial purposes. This paper provide a way of streaming of twitter data and automatically classify these twitter data into different categories using naïve-bayes classifier and Hadoop which perform processing of large data.

Keywords: *Sentiment Analysis, Streaming API's, Hadoop, Naïve-Bayes Model.*

Introduction

Twitter is an online social networking service that enables users to send and read short 140-character messages called “tweets”[1]. These tweets sometimes express opinions about different topics, which is of valuable information that can be used to dwell into the thoughts of millions peoples as they are uttering them.

Sentiment analysis refers to various methods of examining and processing data in order to identify a subjective response, it can be used to find patterns in the opinion of the population such as where people are happier or what the public perception about a brand new product is. The focus of our project is to automatically classify each tweet into categories like art, healthcare, household, occurrence and technology.

- Consumers can use sentiment analysis to research products or services before making a purchase.
- Marketers can use to research public opinion of their company and products, or to analyse customer satisfaction.
- Organizations can also use this to gather critical feedback about problems in newly released products.

Related Work

Sentiment analysis is the most popular trend in today's world. Lot of work has been done in this Sector. There has been a lot of research in the area of Sentiment Analysis. Previous research in sentiment analysis like Pang et al. [2] have analysed the performance of different classifiers on movie reviews. The using of enhanced Naïve-Bayes model for fast and accurate sentiment classification which give a faster and more accurate sentiment of data. In this paper Naïve Bayes approach and Hadoop cluster are used for distributed processing of the textual data [3].

The Approach

In this approach we focussed on performing sentiment analysis on big data which is achieved by splitting the various modules of data in the following steps and collaborating with hadoop for mapping it onto different machines part.

A. Tool

The tools which we use in writing this project are eclipse, twitter4j, Hadoop.

B. Real time data collection

The twitter data is necessary for this project and it is obtained from twitter by streaming API's. For the development purpose twitter provides streaming API's which allows the developer an access to 1% of tweets tweeted at that time bases on the particular keyword. The object about which we want to perform sentiment analysis is submitted to the twitter API's which does further mining and provides the tweets related to only that object. Twitter data is generally unstructured. Tweet messages also consist of a timestamp and the user name. This timestamp is useful for guessing the future trend application of our project.

C. Pre-processing and Tokenization of data

The tweets data can have some valuable info about its sentiment and rest of the words may not really help in determining the sentiment. Therefore, it makes sense to preprocess the tweets. I.e.

- **Lower Case** - Convert the tweets to lower case.

- **URLs** - I don't intend to follow the short urls and determine the content of the site, so we can eliminate all of these URLs via regular expression matching or replace with generic word URL.
- **@username** - we can eliminate "@username" via regex matching or replace it with generic word AT_USER.
- **#hashtag** - hash tags can give us some useful information, so it is useful to replace them with the exact same word without the hash. E.g. #nike replaced with 'nike'.
- **Punctuations and additional white spaces** - remove punctuation at the start and ending of the tweets. E.g: ' the day is beautiful' replaced with 'the day is beautiful'. It is also helpful to replace multiple whitespaces with a single whitespace
- **Tokenization** - each word in the sentence is being split into token. E.g: "the day is beautifull" is being tokenised to "the", "day", "is" and "beautiful".
- **Stop words** - a, is, the, with etc. The full list of stop words can be found at Stop Word List. These words don't indicate any sentiment and can be removed.

D. Developing Classifier

The classifier will automatically classify the tweets according to their categories. For developing classifier the training data set is required. Training data set is required to test the accuracy of the classifier. Training data set are set of data with known classification. Another requirement for developing of classifier is the feature extraction. Feature extraction is the way of extracting particular words or data which define the meaning of the particular sentence and can be used to classify the previously unseen data. To test that the classifier give more accurate results to sentiment classification a confusion matrix is introduced into the training data

Table 16.1: Tweets description

Class	Description
Art	Movies, books, painting,
Fashion	Shoes, dress, jewelry, cosmetics
Health Care	Medical related items
House Hold	Home appliances
Occurrence	Deals, offers, discounts on any items
Technology	Electronic gadgets

Here we classify the tweets as follows:

A. Confusion Matrix

It is a specific table layout that allows visualization of the performance of an algorithm. A confusion matrix contains information about actual and predicted classifications done by a classification system. Performance of such systems is commonly evaluated using the data in the matrix. E.g. If the classifier had already been trained. A confusion matrix will summarize the result of testing the classifier for further inspection.

B. Classification

The Bayesian classification is used as a probabilistic learning method (Naive Bayes text classification). Naive Bayes classifiers are among the most successful known algorithms for learning to classify text documents.

$$P(C|m) = P(C) \prod_{i=1}^n P(f_i|C)$$

Where C is the class, m is the twitter message and f is a feature vector. Feature vector are which contain meaning or give some sentiment value to the tweets.

Results

Confusion Matrix of the classifier

Correctly Classified Instances : 390
Incorrectly Classified Instances : 8

Table 16.2: Confusion Matrix

Art	Fashion	Health Care	House Hold	Occurrence	Technology	Total
55	0	0	0	0	0	55
0	87	0	0	0	0	87
0	0	26	0	0	0	26
4	0	0	33	0	1	38
1	0	0	0	57	1	59
0	0	1	0	0	132	133

Accuracy: 97.9899%

Total Classified Instances : 398

Table 16.3: Total number of tweets.

Art	Fashion	HealthCare	House Hold	Occurrence	Technology
41041	30158	19386	8028	24502	35631

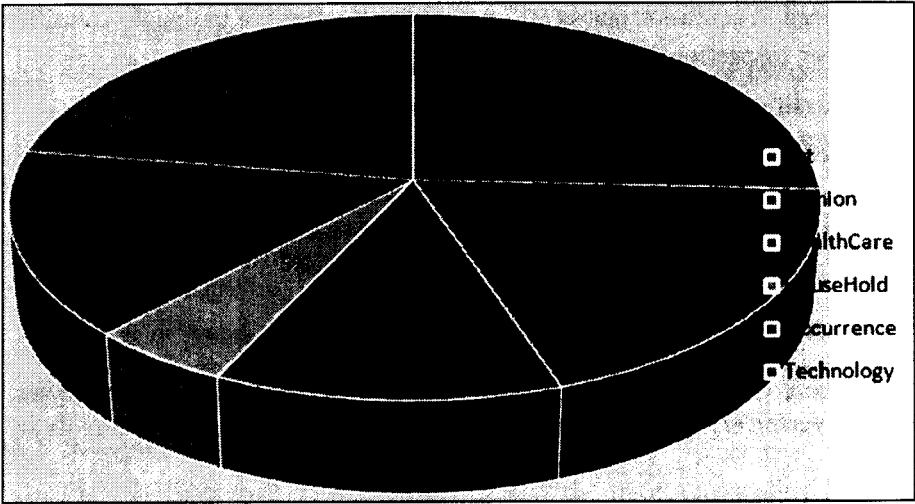


Fig. 16.1: Percentage of tweets

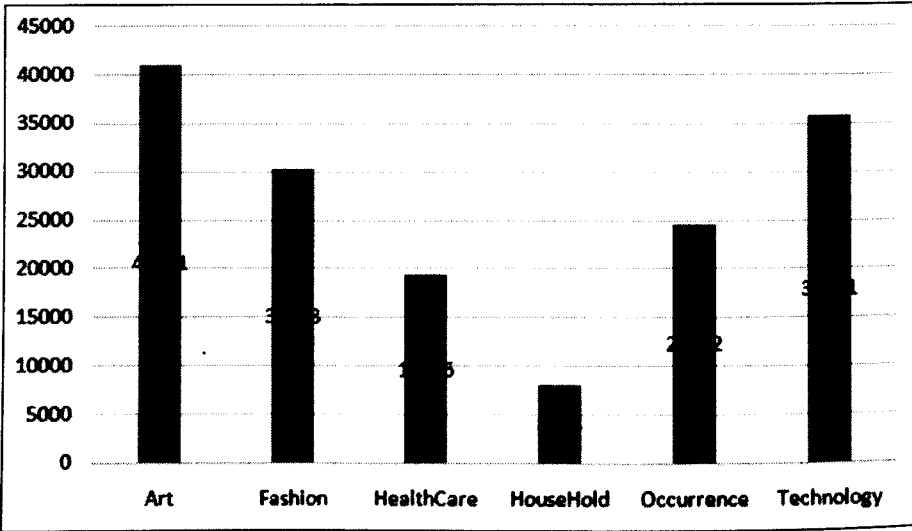


Fig. 16.2: Total numbers of tweets which fall on different categories.

Table 16.3: Most occurrence item

Class	Item 1	Item2
Art	book	Kindle
Fashion	Men's	Shoes
Health Care	Nature	Care
House Hold	Dining	Kitchen
Occurrence	Deals	Hotels
Technology	Sony	Nikon

Classifier of tweets

The above figure shows the result of the tweets message after complete processing and analyzing. They are put into different categories, with art contain the highest and household the lowest number of tweets.

Conclusion

In this paper we have proposed the way of automatic classification of twitter data using Naïve-Bayes approach, which give the accuracy of 98.9899%, and confusion matrix is the way to improve the performance of the Classifier. Hence from the above results we can see that most people enjoy to tweets about art, entertainment and very less people are interest in tweeting about household, and some of the most occurrence item are given the table above. We can use this project also for other social networking site, and for more accuracy of classification in the future we can use some more algorithms like the SVM, Maxent etc.

Acknowledgements

One of the authors, Stevenson Mawa, wished to express special thanks of gratitude to his teacher D. Thabah as well as the Principal of the College, Dr. K. D. Ramsiej of Shillong College for facilitating the golden opportunity to do this wonderful project on the topic Sentiment analysis of twitter data base on hashtag.

References

1. B. Pang, L. Lee, and S. Vaithyanathan, Thumbs up? Sentiment classification using machine learning techniques. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 79{86, 2002}.
2. Vivek Narayanan, Ishan Arora and Arjun Bhatia, "Fast and accurate sentiment classification using an enhanced Naive Bayes model".
3. GVinodhini, R. M. Chandrasekaran, Sentiment Analysis and Opinion Mining: A

- Survey, *International Journal of Advanced Research in Computer Science and Software Engineering*, Page 282-292, 2012.
4. Christopher Scaffidi, Kevin Bierhoff, Eric Chang, Mikhael Felker, Herman Ng and Chun Jin, "Red Opal: product-feature scoring from reviews", *Proceedings of 8th ACM Conference on Electronic Commerce*, pp. 182-191, New York, 2007.
 5. Gang Li and Fei Liu, "A Clustering-based Approach on Sentiment Analysis", 2010, 978-1-4244-6793-8/10 ©2010 IEEE.
 6. Hu, Liu and Junsheng Cheng, "Opinion observer: analyzing and comparing opinions on the Web", *Proceedings of 14th international Conference on WorldWideWeb*, pp. 342-351, Chiba, Japan, 2005.
 7. Jin-Cheon Na, Christopher Khoo and Paul Horng Jyh Wu, "Use of negation phrases in automatic sentiment classification of product reviews", *Library Collections, Acquisitions and Technical Services*, 29 (2005) 180–191.
 8. Go, Lei Huang and Richa Bhayani, "Twitter Sentiment Analysis", *Project Report, Stanford*, 2009.

Index

- Ad-hoc networks, 46, 54, 56, 82
- Ad-hoc On Demand Distance Vector routing protocol (AODV), 47
- Amazon Cloud, 1
- AODV, 45-47, 51-56, 59-64, 66, 68-80
- Asymmetric Encryption, 31
- Background Traffic, 112
- Bandwidth utilization, 82
- Batch processing Applications., 112
- Botnets, 28
- Bridged Network, 99
- Busty Traffic, 111-112
- BX-Book-Ratings, 133
- CapEx, 11
- Certification revocation List, 113
- Cloud Architecture, 4, 15, 121
- Cloud Computing, 1-12, 15-19, 23, 32, 91-98, 110-111, 115-117, 119-120, 122, 124-126
- Cloud data centres, 120
- Cloud Infrastructure Components, 5
- Co-residency Detection, 19, 23
- Cold Start Problem, 127-130, 135
- Collaborative Filtering, 127-129, 135-137
- Collision, 31, 58, 67, 69, 71, 73, 75, 77, 79
- Cryptography, 24, 26-27, 31, 39, 43, 79, 147, 150-151
- Cyber crime, 85, 87-88
- Data disclosure risk, 146, 153, 156
- Data utility, 146, 153, 156
- DHCP, 99, 102, 104-106, 108
- DNS traffic, 113
- Dynamic Topology, 82
- E- mail, 112
- Encryption, 20-21, 25, 27-28, 30-31, 33, 88, 122, 124-125
- End to end delay, 45, 50, 53
- Fisheye State Routing Protocol (FSR), 49
- Forecasting, 138-140, 143-145
- FP Rate, 134
- FTP, 42, 111, 126
- Fuzzy Logical Relationship, 138, 140, 142-143
- Fuzzy Time Series, 138-140, 143-145
- Gartner's findings, 10
- Google Cloud, 1
- Grid computing, 3-4
- Hacking tools/techniques, 85
- Hadoop, 158-159
- HadoopMapReduce, 16
- Hashing, 31
- Hijacking, 110-111, 113-118
- HTTP, 13-14, 33, 42, 44, 80, 87-88, 90, 97-98, 109, 111, 117-118, 126, 136-137, 157
- Hybrid Cloud, 9, 92, 123-124

- Hypervisors, 99-100
- IaaS, 6-10, 13-14, 16, 19, 91-92, 94, 97, 123
- Identity Privacy, 34-36, 38-42
- Information Loss Measure, 153
- Information security, 26, 43, 85-86
- Insider threat, 29
- Interactive Traffic, 112
- Internet, 1-2, 4-5, 7, 9, 15, 17, 25-26, 29, 31-32, 36, 54, 56-57, 59, 85-87, 89-92, 97, 106, 111-115, 117, 120-122, 124-125, 127-128
- Internet protocol like news (NNTP), 112
- IP Address, 48, 99, 103-105, 107-108
- Jitter, 45, 50, 53
- k-anonymity, 147-148, 156
- Kerberos, 113
- Latency Sensitive Traffic, 112
- Layer Two Tunneling Protocol (L2TP), 122, 125
- LDAP (Light weighted Accesses Protocol), 112
- MAC Address, 99, 102-104, 107-108
- Man-in-middle attack, 117
- MDAV2k method, 146-147, 150-152, 154
- Micro-aggregation Techniques, 147-148, 156
- Microaggregation, 146, 156
- Microsoft Azure, 1
- Mobile Network, 34, 36, 44, 56-57
- Mobility, 29, 35-40, 42, 44-46, 48-49, 51, 53-54, 56-58, 82, 96, 120
- Model, 1, 7, 10-11, 15-17, 19, 24, 45-46, 53, 92, 96-97, 123, 126, 128-130, 136, 139-140, 144-145, 153, 157-159, 163
- Multi-tenancy, 17-18, 120
- Naïve-Bayes Model, 158-159
- National Institute of Standards and Technology, 1
- NCTUns-6.0, 55, 66
- Network Operator, 34, 41-42, 122
- Neural Network, 55, 58, 61-66, 79
- Node cooperation, 83
- Non- Real time traffic, 112
- Non-negative Matrix Factorisation, 128
- Online gaming, 112
- Online purchasing, 112
- Online safety, 85
- OpEx, 11
- Outsourcing, 10, 17
- PaaS, 6-10, 13-14, 16, 91-92, 94
- Packet Drop, 66-67, 69, 71, 73, 75, 77, 79
- Password management, 28
- PDR, 45
- Privacy laws, 29
- Private Cloud, 9, 92, 94, 96, 121-124
- Provable Data possession, 20, 23-24
- Public Cloud, 9, 92, 94-95, 120-121, 123-125
- Quasi-identifier, 148
- RADIUS (remote authentication dial in user service), 113
- Random Way point, 46, 53
- Random Waypoint Mobility, 45-46, 51
- RDP (Remote desktop protocol), 113
- Recommender Systems, 127-129, 135-137
- Routing Protocol, 46-49, 54-56, 58-60, 62-63, 66, 79-80, 84
- SaaS, 6-9, 13-14, 16, 91-92, 94-95, 97, 123
- Sabotage, 28
- Security Challenges, 17
- Security principles, 26
- Self-service Cloud Computing, 19, 23
- Semi-trusted third party protocol, 146-147, 150, 154

- Sentiment Analysis, 158-159, 163-164
- Signal-Hiding Techniques, 27
- SMB, 91-92, 97, 117
- SME, 91-92, 96-98
- Social networking site, 86, 163
- Spanning, 99, 104
- SSL transaction, 112
- Streaming APIs, 158-159
- SUSE Linux Enterprise Server, 99, 102
- Symmetric Encryption, 21, 30-31
- Threats to Cloud Integrity, 20
- Throughput, 45, 49, 51, 58, 66-67, 70, 72, 74, 76, 78-79, 81, 83, 92
- Time-Invariant, 138-140
- TP Rate, 134
- Traffic, 28, 48-51, 56-57, 59-60, 66, 83, 95, 110-113, 116-117
- Traffic Hijacking, 110-111, 116-117
- Traffic Protocols, 112
- Tree Algorithm, 99, 104
- Trusted cloud computing platform, 19
- Twitter, 128, 158-159, 161, 163-164
- Universal Mobile Telecommunication System, 37
- VANET, 55-62, 66, 79-80
- Vehicular Ad-hoc Networks (VANETs), 56
- Vertically partitioned data, 146-148, 150, 154, 157
- Video conferencing, 112
- VMware Workstation, 99-104, 107
- VoIP, 112
- Web browsing, 112
- Wireless network security, 27
- Wireless sensor networks, 45, 50, 53-54



The Book

This book is the result of a National Seminar on "Emerging Trends in Advanced Networking and Cloud Computing". This publication contains the papers presented by various scientists, academicians and eminent personalities in the Seminar. The papers depict the emerging trends and some latest developments along with their potential uses in the field of computer science, networking and data mining. With newer developments, this area of technology will continue to progress further and information must reach to every stakeholder for the safe and secure use of the technology with public welfare as the supreme goal,



The Editors

Dr. M.N. Bhattacharjee is an Associate Professor and Head of the Department of Chemistry, Shillong College. He is involved in Science popularisation programmes and has number of research publications.

Smt. A.M. Mitri is an Assistant Professor and Head of the Department of Computer Science and Applications, Shillong College. She is involved in Computer Literacy programmes, IT related activities and many others.

Banteilang Mukhim is an Assistant Professor in the Department of Computer Science and Applications, Shillong College. He is an MCA from University of Madras, proficient in organising Seminars etc., involved in Computer Literacy programmes, IT related activities and many others.

ISBN: 978-93-83252-80-0



2016

₹ 750/-



EBH Publishers (India)

an imprint of Eastern Book House™

136, M.N. Road, Panbazar, Guwahati-781001