



LAKSH  
Estd. 1996

# IPEM JOURNAL OF COMPUTER APPLICATION & RESEARCH

ISSN No. 2581 - 5571  
RNI NO. UPENG/2017/72946

VOL. 2 JUNE 2018

The Annual Refereed Journal of the Computer & IT Department of the  
Institute of Professional Excellence and Management

RS. 300 (ANNUAL SUBSCRIPTION)

## C O N T E N T S

**Energy Efficiency Of Cloud Computing: A Review of The Recent Advances**

*J.K. Seth, Anmol Vashistha & Rahul Mittal*

**A Review on Identification of Fake Profiles in Online Social Network**

*Deepika Varshney, Unnati Singh & Shubh Saxena*

**A Step Towards Protecting Innocent Civilian of Internet Terrene**

*Anvita awasthi & Harshit Sharma*

**Security For Sap Erp System - A Case Study**

*Dr. C.S. Azad*

**Movie Recommender System By Genetic Algorithm**

*Manoj Kumar, Umesh Kumar Patel & Mansi Gupta*

**Image Classification Through Convolutional Neural Network**

*Anupama Sharma, Shivam Dhama & Vishal Rana*

**Big Data & Hadoop: Characteristics And Importance**

*Gaurav Tyagi & Gaurav*

**Information Security & Cyber Crime: An Inside Story of threats, lacking & solution**

*Tushar Sharma*

**Use Of Augmented Reality To Build Various Application Making World More Interactive and Informative**

*Priyank Mathur & Harshit Kumar*

**Internet Protocol Ipv5 Problems & Solutions Over Ipv6**

*Sameer Asthana & Khushboo*



Published by :

**Computer & IT Department  
Institute of Professional  
Excellence and Management**

(ISO 9001:2015 Certified, NAAC Accredited)

A-13/1 South Side G.T Road, Industrial Area  
NH-24 Bye Pass, Ghaziabad (U.P) -201010  
Ph.: 0120-4174500, Fax: 0120-4174500  
E-Mail: journalit@ipemgzb.ac.in  
Website : www.ipemgzb.ac.in

# IPEM JOURNAL OF COMPUTER APPLICATION & RESEARCH

The Annual Refereed Journal of Computer & IT Department of the Institute of Professional Excellence and Management

Vol. 2 June 2018



**Founder, IPEM Group of Institutions**

**Dr. B.S. Goel**

(04.08.1937-10.01.2017)

*A Visionary, Educationist & Philanthropist with values*



★★ SHIKSHA  
BHARATI  
AWARD

*presented to*  
**Dr. B.S. Goel**  
*Executive Director*



★★ INDIAN  
ACHIEVERS  
AWARD

*Best Academic  
Excellence  
Institution in NCR*

## Editorial Board:

- Editor-in-Chief : Dr. R.P.S. Tomar, Director, Department of Computer Science, IPEM College, Ghaziabad, Affiliated to AKTU, Lucknow, rpstomar@ipemgzb.ac.in
- Editor : Dr. Sugandha Goel, Dean Academics, IPEM College, Ghaziabad, Affiliated to AKTU, Lucknow, sugandha.goel@ipemgzb.ac.in
- Associate Editor : Ms. Savita Singh, Assistant professor, Department of Computer Science, IPEM College, Ghaziabad, Affiliated to AKTU, Lucknow, savita.singh@ipemgzb.ac.in
- Associate Editor : Mr. Abhishek Singh, Assistant professor, Department of Computer Science, IPEM College, Ghaziabad, Affiliated to AKTU, Lucknow, abhishek.singh@ipemgzb.ac.in
- Associate Editor : Ms. Mansi Gupta, Assistant professor, Department of Computer Science, IPEM College, Ghaziabad, Affiliated to AKTU, Lucknow, jv.mansi@gmail.com

## EDITORIAL REVIEW BOARD

- Dr. Vineet Kansal : Director, ITS Engineering College, Gr. Noida, Affiliated to AKTU, Lucknow, dir.engg@its.edu.in
- Dr. Sunil Kr. Khatri : Director, Amity School of Information Technology Computer, Amity University, Noida, skkhatri@amity.edu
- Dr. Anil Solanki : Director, BIET, Jhansi, solankibiet13@gmail.com
- Dr. Mridul Gupta : Professor of Computer Science, CCS University Meerut, mkgupta\_2002@hotmail.com
- Dr. Naresh Chauhan : Prof. & Head, Department of Computer Science, YMCA University of Science & Technology, Faridabad, nareshchauhan19@yahoo.com
- Mr. D. N. Sahay (ITS) : General Manager (Satellite), Satellite Building, Advanced Level Telecommunication Training Centre (ALTTC), Ghaziabad, Government of India, devanandsahay@gmail.com
- Dr. Devendra Tayal : Associate Professor & Head Indra Gandhi Institute of Technology (GGSIU, New Delhi), dev\_tayal2001@yahoo.com

Printed and Published by Mr. Anupam Goel on behalf of Laksh Educational Society and Printed at Ghaziabad Offset Press, 133, East Model Town Tehasil Road, Ghaziabad and Published at Institute of Professional Excellence and Management, A-13/1, South Side G.T. Road Industrial Area, NH-24 Bypass Ghaziabad (U.P.) 201010 -INDIA.  
Editor: Dr. Sugandha Goel

All rights reserved. No part of this publication may be reproduced in any form or by any means, electronic, photocopying or otherwise, without written permission of managing Editor, Journal of Computer-Application & Research.

# Editor-in-Chief's Message

GREETINGS!

It is with great pride, eagerness and presentiment to read second issue of the series of Journal, "IPEM JOURNAL OF COMPUTER APPLICATION & RESEARCH". The Journal consists the research papers of three distinct and mutually exclusive disciplines of knowledge like Advanced Computing, Networking Informatics & Security. However, convergence of these technologies is observed in many real world applications, including cyber-security, internet banking, healthcare, sensor networks, cognitive radio, pervasive computing amidst many others. The Journal also includes innovative computing techniques and relevant research papers in informatics with selective applications in pattern recognition, signal/image processing and HCI



The application of technology is a key theme in every paper which is published in this journal. The intent of this journal is to showcase technologies which could bring about a fundamental change in achieving Researchers impact.

The present issue has been very carefully put together covering a range of Advanced Computer Application, Networking Technologies & Security. The contributions have come from Academics from renowned Institutions, Research Scholars, Students and IPEM faculty.

As the Chief Editor of " IPEM JOURNAL OF COMPUTER APPLICATION & RESEARCH" I would like to take opportunity to express my sincere gratitude to the authors who have chosen this Journal to disseminate their research work. Further I would like to thank Mr Anupam Goel Secretary Laksh Educational Society, Director General & Dean Academics of IPEM, the readers, the content providers, who have made this Journal the best possible

A great Journal cannot be made great without a dedicated Editorial Team of Editors and reviewers. I would also like to thank the Editorial Team Members, Review Committee Members, printer for the success of this Journal.

We look forward to receive the contribution for our next issue from Academicians, Scholars and Professionals to ensure consistency and the success of Journal. We welcome comments and suggestions that would further enrich this Journal.

I am happy to wish that our Journal would experience steady and healthy growth

**Dr. R.P.S. Tomar**  
Editor-in-Chief

# From the Editorial Board

**W**e are glad to present the second edition of the Computer and IT Department journal "IPEM JOURNAL OF COMPUTER APPLICATION & RESEARCH" JUNE 2018. Publishing a research journal is a tough task.

However, we braved all the odds, and published the issue as always, on time. We followed a rigorous method to select the papers. All the papers we have included in this issue of IPEM JOURNAL OF COMPUTER APPLICATION & RESEARCH are peer reviewed and only those papers which went through this rigor have been given space in this journal.

This journal attempts to document and spark a debate on the research focused on technology in the context of emerging technologies. The area could range from Big Data and Hadoop, Cloud Computing, Cyber Crime, Internet Protocols, Neural Network etc. These technologies could be very sophisticated to very elementary but in term of impact they would be capable of being commercialized, scaled up and focus on real life challenges.

We sincerely hope that these in-depth research papers, focusing on different issues, will further stimulate the academic research, and will help in developing an insight in the concerned areas. We are eagerly waiting for your critical response which we shall incorporate in the forthcoming issues. We are greatly indebted to the paper writers who took keen interest and submitted their research paper on time. It is because of the sincere efforts of these peoples that the IPEM JOURNAL OF COMPUTER APPLICATION & RESEARCH is in your hand today.

We are grateful to our secretary Mr. Anupam Goel who provided all the moral and financial support to publish the IPEM JOURNAL OF COMPUTER APPLICATION & RESEARCH.

# IPEM JOURNAL OF COMPUTER APPLICATION & RESEARCH

The Annual Refereed Journal of Computer & IT Department of the Institute of Professional Excellence and Management

Vol. 2 June 2018

## Contents

- 01 **Energy Efficiency Of Cloud Computing:  
A Review of The Recent Advances** 01  
*J.K. Seth, Anmol Vashistha & Rahul Mittal*
- 02 **A Review on Identification of Fake Profiles  
in Online Social Network** 08  
*Deepika Varshney, Unnati Singh & Shubh Saxena*
- 03 **A Step Towards Protecting Innocent  
Civilian of Internet Terrene** 13  
*Anvita awasthi & Harshit Sharma*
- 04 **Security For Sap Erp System - A Case Study** 17  
*Dr. C.S. Azad*
- 05 **Movie Recommender System By Genetic Algorithm** 22  
*Manoj Kumar, Umesh Kumar Patel & Mansi Gupta*
- 06 **Image Classification Through Convolutional Neural Network** 31  
*Anupama Sharma, Shivam Dhama & Vishal Rana*
- 07 **Big Data & Hadoop: Characteristics And Importance** 36  
*Gaurav Tyagi & Gaurav*
- 08 **Information Security & Cyber Crime: An Inside Story of  
threats, lacking & solution** 40  
*Tushar Sharma*
- 09 **Use Of Augmented Reality To Build Various  
Application Making World More Interactive  
and Informative** 46  
*Priyank Mathur & Harshit Kumar*
- 10 **Internet Protocol Ipv5 Problems & Solutions Over Ipv6** 50  
*Sameer Asthana & Khushboo*





# Energy Efficiency of Cloud Computing: A review of the Recent Advances

J.K. Seth\*  
Anmol Vashistha\*\*  
Rahul Mittal\*\*\*

## ABSTRACT

*This paper will analyse and review the latest and the most recent advances in the techniques that are in use and are being developed to make cloud computing much more energy efficient. Along with this it will also feature the details about various components that come together to make up the cloud, what are the energy consumptions of these components. There will also be an analysis of the scope of making entire cloud energy efficient, by reviewing the available algorithms, scheduling techniques, management policies and allocation method for the resources.*

**Keywords:** Information and Communication Technology (ICT), Service-oriented Architecture (SOA), Virtual Machine (VM), Solid State Drive (SSD), Cloud computing.

## Introduction

Cloud computing has completely transformed entire Information and Communication Technology (ICT) industry since its emergence. Cloud computing is both, a business as well as an economic model which has gained immense popularity due to its unique ability to realize hardware and software as products. It has paved a way for the enterprises to reduce hardware and software cost, along with reduction of maintenance and support staff. Data centers and cloud computing services providers, have made sure that adoption of cloud computing technology brings enormous profits and revenue to both the adopters and themselves. Though this technology has a great number of advantages, but it doesn't come without its fair share of controversies. The biggest and the most debated of which is the Energy Efficiency of Cloud Computing. Cloud computing is a cost-effective and highly scalable infrastructure for running HPC, enterprise and Web applications. Hence, to design such solutions, deep analysis of Cloud is required in context of their power efficiency. Before proceeding further, firstly we need to understand more about cloud computing

and the various components that come together to form the Cloud.

## Cloud Computing

Cloud computing provide various computer processing data & resources whenever they are required. Cloud computing has originated through the evolution and adoption of existing technologies and paradigms. The primary objective of cloud computing is to enable users to take benefits from all the available technologies, without the need for in-depth knowledge or expertise about all of the components. The primary motivation behind cloud is to reduce costs, and helps the users focus on their core business ideas instead of being impeded by IT obstacles. The main technology responsible for enabling cloud computing is virtualization. With operating system-level virtualization it becomes easy to create a scalable system of multiple independent computing devices, and to more effectively use computing resources that were idle resources by allocating them efficiently. Autonomic computing automates this process, through which the user can provision resources as per his demands. By reducing user involvement, automation speeds

\*Asst. Professor, Ajay Kumar Garg Engineering College, Ghaziabad, [mrjkseth@gmail.com](mailto:mrjkseth@gmail.com)

\*\*Student, Ajay Kumar Garg Engineering College, Ghaziabad, [anmolvashisthaav@gmail.com](mailto:anmolvashisthaav@gmail.com)

\*\*\*Student, Ajay Kumar Garg Engineering College, Ghaziabad, [rahulmittal312@gmail.com](mailto:rahulmittal312@gmail.com)

up the process, minimizes labor costs and reduces the possibility of human errors. Users routinely face difficult business problems.

Cloud computing is based on concepts of Service-oriented Architecture (SOA) which helps the user to

break these problems into services that can be integrated to give a solution. Cloud computing provides the entirety of its resources as services, and makes use of the well-established standards and best methods available to allow global and easy access to cloud services in an easy and efficient manner. [1]

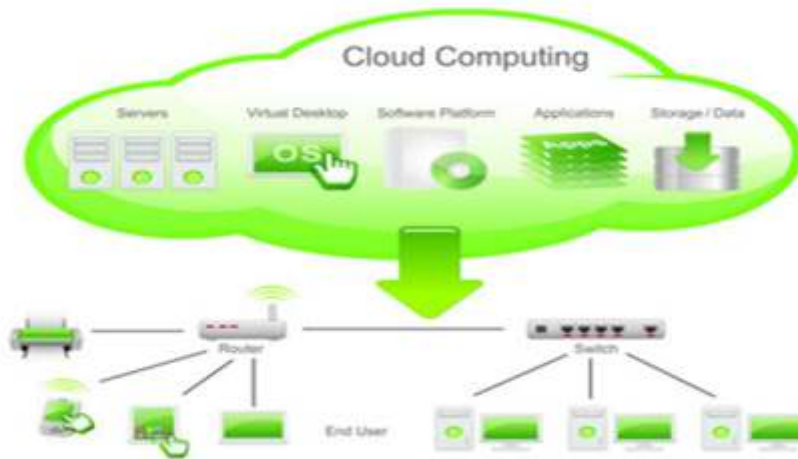


Figure 1: Cloud Computing Architecture

## 2.1 Green Cloud Computing

Green Cloud computing is the eco-friendly and use of the environmentally responsible Cloud Architecture and its resources. In broader terms, it is also defined as the study of designing, manufacturing/engineering, using of cloud resources in a way that reduces their environmental impact as well as makes them much more energy efficient. It can also be defined as using Internet computing services or cloud-based services from a service provider that has taken measures to reduce their environmental impact. In general, it's cloud computing with less environmental impact than 'normal' cloud computing.

## 2.2 Energy Efficiency

Energy efficiency refers to a minimization of energy used for a given service or level of activity, as defined by the World Energy Council. However, defining the energy efficiency of typical data center equipment is extremely difficult because it represents a complex system with a large number of components from various research areas such as computing, networking, management, and the like. Energy waste refers to energy used by the systems' main task but without useful output (e.g., energy used while running in idle mode)[2].

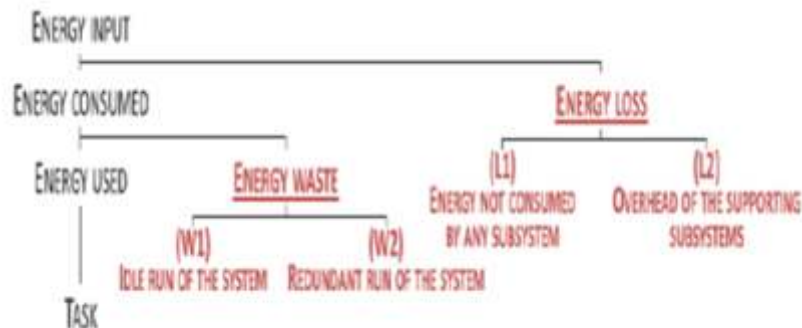


Figure 2 Critical points within a system where energy is lost or wasted



### 2.3 Service Based Layers in Cloud Computing

Cloud computing is mainly made up of three layers which cover the entire computing stack of a system. Each of these layers provide different set of services to end users. At the lowest layer, Cloud services are named as Infrastructure-as-a-Service (IaaS) which is composed of virtual machines or physical machines, storage, and clusters. Cloud computing can likewise be heterogeneous, as they may incorporate coordinated bunches, PCs and workstations. Besides, the framework foundation can likewise incorporate database administration frameworks and other stockpiling administrations. The framework is overseen by an upper administration layer that gives runtime condition customization, application separation, bookkeeping and nature of administration. The virtualization apparatuses, for example, hypervisors, additionally sit in this layer to deal with the asset pool and to parcel physical framework as tweaked virtual machines. Contingent upon the end client needs, the virtualized foundation is pre-arranged with capacity and programming condition, what spares time for clients who don't have to fabricate their framework sans preparation. [1]

- **Software as a service (SaaS)**- Its is something which charged for the Time of using application. It is widely used by the end users , which have minimum usage. Just like google and Microsoft are providing web services to the users , so they can access them without need of installing it.
- **Platform as a service (PaaS)**- In PaaS, a service provider facilitates services to the users with a set of software programs that can solve the specific tasks. Lets take an example of Java compiler or C/ C++ compiler, If you are a

developer or coder , so you must have complies with you all time, but cloud services makes it easier for you, by providing any time any where compilation facility. Toady you just need to write code , and you can compile it online to generate output.

**3. Infrastructure as a service (IaaS)**- It is a customised services, which is built in operating systems that are already running. IaaS means you can rent a server and pay according to used resources. The cloud service provider facilitates services to the users with virtual machines and storage to improve their business capabilities. Its target audience is software developer.

### Components of Cloud Computing and their Energy Efficiency

Cloud computing can be considered as a system with three main components.

#### i. Data Centers

Data centers are the heart of ICT and highly essential for the function of communications, business, academic, and government systems. But data centers have been causing concerns because of their high energy consumption. Due to the high demand for data processing and storage, many data centers have been built in the past. The environmental impact of this high energy consumption is also causing concern because the carbon footprint of ICT has been growing rapidly. There have been improvements in the energy efficiency of data centers due to initiatives by governments, organizations, and professionals to tackle the problem. In 2007, Google and Intel started the

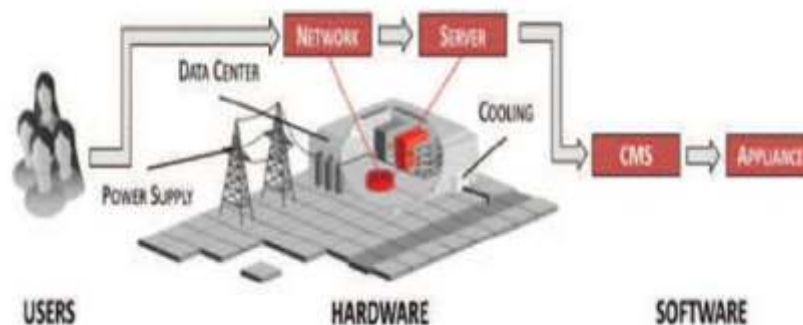


Figure 3: Domains of a typical Cloud Computing Data Center

Climate Savers Computing Initiative, an organization committed to making computers and servers more energy efficient. The general composition of Data Centers has been shown in the figure 3.

## ii. Communication Networks

Communication networks form the backbone of the cloud computing system. The Internet and telecommunication networks are the links between cloud users and service providers. Cellular and fixed broadband networks are used to transport data back and forth between the data centers and cloud users. Users can only access the cloud computing services, if they have internet access, since these services are based on remote servers in data centers. The energy efficiency of the internet has recently faced intense scrutiny. The exponential growth of the Internet led to the need for corresponding growth of equipment required to route the increased traffic. The consequence of this is growth in energy consumption, the unprecedented expansion of wireless and wired networks has resulted in high increase in energy consumption and significant environmental footprint. This means that measuring the energy efficiency of cloud computing cannot be based only on data centers but also on the networks that truly enable the provision of cloud services. [3]

## iii. User Devices

Cloud user devices are described as clients and classified as mobile devices such as PDAs and Smartphones. Thin clients are the computers without internal hard drives that depend on the servers in the "cloud" to do all computations and display the information for users. Whereas, Thick clients are normal computers that use web browsers to access cloud services. Although battery life has been a concern for mobile device users, the energy consumption of mobile phones is pretty low. Mobile handsets account for only 1% of the total energy consumption of cellular networks. However, for other devices like PCs, since resources are accessed through Internet, both applications and data are needed to be transferred through the compute node. Therefore, it requires much more data communication bandwidth between user's PC to the Cloud resources, with respect to its application

execution requirements. In some cases, if data is really large, then it may turn out to be cheaper and more environmentally sustainable, efficient to send the data by mail, rather than to transfer through Internet. In the network infrastructure, the energy consumption depends especially on the awareness of wired network and its power efficiency, namely the network equipment or system design, topology design, and network protocol design. Most of the energy in network devices is wasted because they are designed to handle worst case scenario. Therefore, the energy consumption of these devices remains almost the same during both the idle state and peak state. Thus, many improvements are required to get high energy efficiency in these devices. [3]

## 4. Resolution for Energy Efficiency Crisis

A number of solutions are discussed in literature few of them are mentioned in this section.

### A. Schedulers

#### i. What is a scheduler?

A scheduler is based on the requests for a specific resource, as well as many other details, which are sent by the customer to the cloud. When such requests are received, a scheduler scans for the resources available, to satisfy the request and to assign them to the user. In order to calculate the best possible placement of the request, the scheduler mostly has information on the current state of the current environment of the concerned cloud along with the history of the cloud. These factors can be used for the creation of the scheduler, and to allow for it to optimise the placement based upon the information often allowing for more efficient schedulers. The aim of the scheduler is to provide true and effective consolidation, which is not as simple as keeping servers at 100% utilisation.

#### ii. Scheduling Algorithms for better energy efficiency

Green Clonal Scheduling Optimization Algorithm (GCSOA): GCSOA algorithm is based on the clonal selection algorithm that is used by the natural immune system to define the basic features of an immune response to an antigenic stimulus. Experimental results show that this algorithm can

not only effectively minimize the execution time and energy consumption, but can also achieve efficient resource load balancing, thus effectively improving the resource utilization and scheduling efficiency leading to a better and much more energy efficient task scheduling. [5]

### B. VM management policies

In addition to the use of schedulers to control the requests, proper management of the requests can be done in real time to help optimize the use of the servers. These policies follow the approach which states that the VM allocation policy is used to calculate utilization thresholds for the servers within the cloud. Using the selection policy, VMs are selected for migration from the server, when and if a server is above or below the specified values. When the server is not in use then it will be put into a low power state to save energy. [4]

#### i. The First Fit Decreasing algorithm

The first algorithm, is a modification of the greedy algorithm described in the literature [7],[10], called First Fit Decreasing. This algorithm sorts the VMs on the basis of the list of requests that it gets, into a non-increasing order by CPU usage and then processes the entire list in that order by placing each VM into the very first host which fits. It then continues to the next VM until all VMs are allocated. Finally, it suspends idle physical hosts in order to reduce power consumption [7]. If no VMs are allocated to a host, it suspends that host, greatly reducing the energy consumption. [5]

#### ii. The Power Aware Best Fit Decreasing algorithm

The second algorithm is based on the works of Beloglazov, called the Modified Best Fit Decreasing algorithm (MBFD) [8] or, Power Aware Best Fit Decreasing algorithm (PABFD) [9]. According to [9], this algorithm is a variation of FFD algorithm. While the algorithm itself is similar to FFD, the main difference in it is that PABFD calculates the increase in power consumption after performing the allocation of the current VM in each host in data center, storing the lowest increase in power consumption. The VM is then allocated to the host which had the lowest increase in power consumption. Finally, it suspends all the idle physical hosts. [5]

### C. Other allocation methods

In addition to the schedulers and VM policies mentioned above, there are other various scheduling and allocation methods that are used within cloud environments. For example, there are multiple different schedulers that are used with the allocation of resources in auction based environments. They can be used in many different scenarios such as assigning a price to a customer for their request or assign a customer to a company, this is possible only if this auction based environment consists of multiple different organizations. Once the auction has concluded, the customer's request is then placed onto a server for its execution. Such an algorithm for a green auction based scheduler has been proposed by Huu and Tham called Green-greedy algorithm. In this algorithm users make a bid for a certain request with a cost and energy value against it, which is then compared to other bids. Once the bids have been accounted for and compared, the user can improve the time for placement, by offering to increase the price or energy of the request. These bids are then processed and sorted in the order of which bid is best on the basis of the energy and price of the request. [6]

## 5. Features of Clouds Enabling Green Computing

Although there is a great concern in the community that Cloud computing will result in higher energy usage by the datacenters, still the technology enabling Cloud computing has a green lining. There are several different technologies and concepts employed by various Cloud providers to achieve better utilization and efficiency with respect to the traditional computing. Therefore, this leads to a lower carbon emission, which persists due to highly energy efficient infrastructure and reduction in the IT infrastructure itself by use of multi-tenancy. The key innovation indispensable for vitality proficient Clouds is "Virtualization," which permits huge change in the general vitality effectiveness of Cloud suppliers by making utilize the economies of scale related with expansive number of associations having a similar framework. By the union of underutilized servers as numerous virtual machines which share same physical server at higher use, organizations can increase substantially higher reserve funds regarding space, administration, and vitality..

According to a report by Accenture [7], following three key factors that have enabled the Cloud computing to lower energy usage and carbon emissions from ICT. Using these Cloud features, organizations can reduce carbon emissions by at least one-third, per user by moving their applications to the Cloud. These savings are driven by the high efficiency of large scale Cloud data centers. These factors are:

### A. Dynamic Provisioning

In the traditional setting, private infrastructure and datacenters used to be maintained just to fulfill worst case demand. Thus, IT companies ended up deploying far more infrastructure than needed for the average use scenario. There are many reasons for such over-provisioning. Firstly, it is very difficult to predict the demand at a time instant which is particularly true for Web applications. Then, it is also difficult to guarantee availability of services as well as maintain certain level of service quality to end users. To handle peak loads during short period in a year, for example, e-commerce sales, running hundreds of server throughout the year is not really energy efficient. Thus, the infrastructure provisioned with a conservative approach in such cases results in unutilized resources. Such situations can be promptly and effectively overseen by Cloud framework. The virtual machines in a Cloud foundation can be live moved to another host, on the off chance that when a client application requires more assets. Cloud suppliers screen and anticipate the request and therefore designate assets as indicated by request. Those applications which require less number of resources can be consolidated on the same server. Thus, datacenters always maintain the servers in active state, according to current demand, which results in an overall state of low energy consumption with respect to the conservative approach of over-provisioning. [1]

### B. Multi-tenancy

Using the approach of multi-tenancy, Cloud computing infrastructure can reduce overall energy usage and associated carbon emissions. The SaaS providers serve various companies on same infrastructure and software. This approach is obviously much more energy efficient in comparison to the approach of keeping multiple

copies of software installed on different infrastructure. Furthermore, most often, the different businesses that avail these cloud services, have highly variable demand patterns, and hence multi-tenancy on the same server enables the flattening of the overall peak demand which reduces the need for extra infrastructure. Smaller the fluctuation in demand better will be the results in prediction and greater energy savings.

### C. Server Utilization

In general, the on-premise infrastructure, runs with very low utilization, sometimes it goes down up to 5 to 10 percent of the average server utilization. Using virtualization technologies, many applications can be hosted and executed on the same server in isolation at the same instant, thus leading to utilization levels of up to 70%. Hence, it dramatically reduces the overall number of active servers. Even though high utilization of servers results in more power consumption, but server running at a higher utilization can process much more workload with similar power usage than the one at a low utilization level. [1]

## 6. Miscellaneous Techniques

### A. Hardware Cooling Tactics

As cloud providers select their hardware including servers, storage and all the networking gear, they must be more closely concerned and keep in mind the factor of power consumption for the cooling of all this hardware that is the power that will be needed to keep those devices cool. It is not enough for cloud providers to look just at the benchmarks for electricity that is required to bear all the IT load. The consumption figures must be studied much more closely.

Another point to consider is that whether the equipment that will be installed, allows for the data center layout and design choices such as using hot and cold aisles that can reduce the cooling load. Cooling costs can account for 30% to 40% of the power load of many data centers, according to Phil Nail, chief technology officer of AISO.net, a hosting provider in Romoland, California, that uses about 200 solar panels to power its operation.

## B. Disciplined Storage Architecture

The storage tiering methodology that enables organizations to make exhaustive information administration procedures can be similarly compelling as a method for lessening power utilization for cloud suppliers. This should be possible by putting chronicled information on slower, bigger drives that utilization less power and holding speedier gadgets for mission-basic data that should be gotten to all the more rapidly in contrast with other info. Service suppliers can oversee capacity spending plans all the more viably, as well as spare a portion of the related vitality costs. By separating the stockpiling into levels it ends up less demanding to put bring down I/O mass stockpiling on slower, huge plates as opposed to utilizing higher-speed circles or SSDs. This will bring about a considerably littler impression in the general vitality utilization for the cloud and will fundamentally lessen control utilization and warmthage.

## 7. Conclusion

From the above review and analysis, we can conclude that although it is believed that the energy consumption of the cloud eclipses its overall benefits, it is just not true. Using various techniques and by making smarter choices the energy efficiency of the Cloud Computing Platforms can be increased to up to two folds. Use of better scheduling algorithms, innovative allocation methods and effective VM management policies like consolidation of VMs, the problem of environment sustainability of cloud can be easily solved. Also, the features already made available by the cloud computing technologies like multi-tenancy and dynamic provisioning should be taken into account and used. Applying better ways of utilizing the servers to their full performance and by using smarter hardware choices along with tiered storage architecture the Cloud Computing can easily

transform into the Green Cloud Computing which takes care of the environment, has much less of a carbon-footprint and is more energy efficient.

## References

- [1] Saurabh Kumar Garg and Rajkumar Buyya, "Cloud computing and Distributed Systems (CLOUDS) Laboratory", Dept. of Computer Science and Software Engineering, The University of Melbourne, Australia.
- [2] Nathan Whittington, Lu Liu, Bo Yuan, Marcello Trovati, Department of Computing and Mathematics, University of Derby, Derby, DE22 1GB, U.K. "The Key Laboratory of Embedded System and Service Computing, Ministry of Education, Tongji University", Shanghai, China.
- [3] Yang Liu, Wanneng Shu, and Chris Zhang, "A Parallel Task Scheduling Optimization Algorithm Based on Clonal Operator in Green Cloud Computing", *Journal of Communications*, Vol. 11, No. 2, pp. 185-191, February 2016.
- [4] K. J. Ye, Z. H. Wu, X. H. Jiang, and C. M. Ho, "Virtualization cloud computing energy management platform," *Chinese Journal of Computers*, vol. 35, no. 6, pp. 1262-1284, June 2012.
- [5] A. J. Hussain, P. Fergus, H. Al-Askar, and D. Al-Jumeily, "Dynamic neural network architecture inspired by the immune algorithm to predict preterm deliveries in pregnant women," *Neurocomputing*, vol. 151, no. 3, pp. 963-974, March 2015.
- [6] Z. H. Zhang, S. G. Yue, M. Liao, and F. Long, "Danger theory based artificial immune system solving dynamic constrained single-objective optimization," *Soft Computing*, vol. 18, no. 1, pp. 185-206, Jan. 2014.
- [7] L. Chiaraviglio and I. Matta, "Green coop: Cooperative green routing with energy-efficient servers," in *Proc. 1st ACM International Conference on Energy-Efficient Computation and Networking, e-Energy*, Beijing, 2010, pp. 191-194.
- [8] W. Wei and Y. Qi, "Information potential fields navigation in wireless ad-hoc sensor networks," *Sensors*, vol. 11, no. 5, pp. 4794-4807, May 2011.



# A Review on Identification of Fake Profiles in Online Social Network

Deepika Varshney\*  
Unnati Singh\*\*  
Shubh Saxena\*\*\*

## ABSTRACT

*In the present scenario, with the inevitable use of social networking sites, the most prominent are the creation of fake identities which has a major contribution in the prolonged threat and venomous activities over the internet. The paper focuses on the research work done in order to detect fake profiles in online social network using various techniques. The ways of detecting fake social media accounts include analyzing individual accounts and capturing their malicious activities spanning a large group of accounts. This paper study various detection techniques to identify false identity and shows the comparison among them with respect to how efficient the respective technique to identify fake profiles. Our survey provides a good understanding of the problem, current solution space, and scope of future research to deal with false identity attacks efficiently.*

**Keyword-** Fake Accounts, Advanced Persistent Threat, Spear Phishing, Spam.

## Introduction

### A. Fake Profile: Definition

We all have a certain image of ourselves beliefs about the kind of person we are. A clear sense of who you are makes your identity. Identity is the concept you develop about yourself that evolves over the course of your life. A typical example is that of your passport which include your name, birth date and place, nationality, photograph and digitally captured fingerprints. Once you submit your identifications, the authorities verifies whether the identity proofs belong to the same person or not. This is how you become a true owner of your passport. In recent years, social networking sites have become a global mass phenomenon. These provide networked communication platform in which users have uniquely identifiable profiles that contains persons name, birth date and place, pictures, likes and dislikes. A statistics published in June 2017 indicates Facebook as the largest online social network with 2.07 billion monthly active users. These sites provide easy accessibility to create

accounts without concerning whether the person creating and controlling the account is real or fake. This is called false identity. False identities play a vital role in Advanced Persistent Threat (APT). An APT is a network attack in which an unauthorized person gains access to a network and stays there undetected for a long period of time. The intention of an APT False identities play a vital role in Advanced Persistent Threat (APT). An APT is a network attack in which an unauthorized person gains access to a network and stays there undetected for a long period of time. The intention of an APT attack is to steal data rather than cause damage to network or organization. False identities are involved in malicious activities like spamming, artificially inflating a number of users to promote an application or to like/comment a page or post etc. Social media has become the most prevalent platform to communicate among today's youth. With this, false identities are using the platform to create fake identity with personal information that looks genuine, to establish trust with the person on target. The target is then exploited by its trustworthy source by directly interacting to obtain information,

\*Assistant Professor, Ajay Kumar Garg Engineering College Ghaziabad, Delhi, India, Email: deepikavarshney06@gmail.com

\*\*Student, Ajay Kumar Garg Engineering College Ghaziabad, Delhi, India, Email: sengar.unnati95@gmail.com

\*\*\*Student, Ajay Kumar Garg Engineering College, Ghaziabad, Delhi, India, Email: ssshubh97@gmail.com



through spear phishing attack. Spear phishing focuses on specific individual within an organization and social media accounts to seek unauthorized access to sensitive information through email spoofing.

These accounts seem to be legitimate but when detected, they turn out to be false. They contain personal information which does not belong to the person who created the account. Such accounts are also termed as faked accounts. The items taken for identifications are certified by the authorities of the country. As a person cannot issue an identity card by its own, different institutions are responsible to provide identifiers to that person. For example, Government provide aadhaar cards as a proof of Indian citizenship, banks and other financial organizations issue credit and debit cards, authority provide passport as identification proof to travel abroad and identity cards with different standards of reliability. Digital identifiers are assigned in form of unique string of characters called as social security number (SSN). This number is unique and used to differentiate people. With all such identification procedures, one can still make an identifier for himself. It is possible by creating a fake email address or a false social media account. There are no restrictions imposed in making an online profile, thus this facility is misused at larger scale nowadays. Kaplan and Heinlein [1] (2010), defined the challenges and opportunities of social media. They defined social media as a group of Internet-based application, that allow the creation of User Generated Content and is build on the foundation of Web 2.0. They also scratched pieces of advice for companies which make use of Social Media. According to Kietzmann et al. [2] (2011), the building block of social media is the user identity. There are many social media sites with their own perspective regarding user identity. Some promote usage of real identification of the user while others remain unconcerned with the identities and also allow their access by nicknames. Douceur et al. [3] (2002) contradict this statement. They argued that for presenting convincingly distinct identities computing environment needs logically central trusted authority which would manage identity information; which seems to be practically impossible. This paper mainly focuses on the literature review of the state-of-the-art research aiming at detecting fake profiles in social media. The

Paper will discuss over the approaches that will target either individual fake accounts and their connections, or on coordinated activities involving such accounts.

## B Problem Statement

In this paper we did comparison among various techniques used to detect fake profiles. Rest of the paper is organized as follows. Techniques to mark fake profiles is elaborated in Section II. Section III compares the various detection techniques based on their advantages and disadvantages. Conclusion and future work are described in Section IV respectively.

## II. Techniques To Mark Fake Profiles

There are various techniques used in current scenario to mark suspicious videos, some of them we are covering in this paper given as follows.

1. Winoing
2. OSB (Orthogonal sparse bigram)
3. SBPH (Sparse Binomial Polynomial Hashing)

For all websites where a user can get benefits like a free user account, there may occur certain malicious activities which cause harm to the user of that network website. It is the point when the profile is being checked whether it is creating a spam or not! Spam can be in sense of comment, likes or same person having more than two accounts for personal benefit! To prevent such spamming on Online Social Networking Sites (OSN), OSN is used to check the individual account to detect spamming (in most of the cases) and thus can operate in a peaceful manner. In order to do this, features are collected and used with different machine learning algorithms which can easily detect spams or the fake account! The spam can be detected in a numerous ways like Naive Bayes classification, K-nearest neighbour classification, Support Vector Machines and many more. Sites like Youtube can identify the spams using spam comment and fake bot generated likes. Similarly, according to the OSN site Facebook, those who post something more than 50 times a day are suspected to be delivering spam content!. In Machine Learning, we work on basically two types of learning - supervised and unsupervised, it becomes essentially important to make two different classes for classifying all the OSN content in two categories, say good or spam. For

example, Winnowing is a good technique which follows Supervised learning which is described in brief below.

**A. Winnow**

Winnow :-It is a simple technique used to classify the text(particularly useful for OSN to check spam comments) whether it is Spam or good. For this, first of all we retrieve all the text from comment, tokenize it and then we process that text according to the specifications in order to get a bunch of features from it. After that, the Winnow comes into play. Features are calculated on the basis of what the word wants to point at that position! By feature extraction, we mean to extract out the number of features, the machine learning program is designed to learn out of each comment or instance. Here we use Supervised Machine Learning algorithm. So we use a bunch of samples for training our machine for making it initially learn whether the given file is spam or good. In order to create samples for calibrating our machine, we set the extensions of the text files for feature extraction as good or spam. The more the training is given to the system, the more classification accuracy score can be achieved and

hence resulting in better classification of message/comment. Christian Siefkes et al.[4] (2014) had used winnow technique in his work in order to get a good accuracy.

**B. OSB AND SBPH**

Chu et al.[5] performed an analysis of fake profiles on Twitter profile in the year 2010. For this, they manually labelled 3000 profiles humans, bot or cyborgs. They worked for the sake of tweet spam classification and they used Bayesian Classifier for it. It comes under Orthogonal Sparse Bigrams(OSB). OSB is a feature extraction structure and it avoids redundancy. It is known as redundancy because it creates successive pattern without any duplication of effort. OSB and SBPH are used in combination to separate the text into classes, namely good or spam. SBPH ,an acronym used in for Sparse Binomial Polynomial Hashing uses binary coded format where it uses 1 in place of where we use a word and 0 where we skip a word. It is known as sparse because we skip either word and takes it. Table I shows the understanding of how hashing is done in SBPH and OSB.

**TABLE I: Understanding the Hashing in SBPH and OSB**

Number	Sparse Binomial Polynomial Hashing (SBPH)	Orthogonal Sparse Bigram (OSB)
1 (1)	now?	
3 (11)	feeling now?	feeling now?
5 (101)	you<skip> now?	you<skip> now?
7 (111)	you feeling now?	
9 (1001)	are<skip><skip> now?	are<skip><skip> now?
11 (1011)	are<skip> feeling now?	
13 (1101)	are you <skip> now?	
15 (1111)	are you feeling now?	
17 (10001)	How <skip><skip><skip> now?	How <skip><skip><skip> now?
19 (10011)	How <skip><skip> feeling now?	
21 (10101)	How <skip> you <skip> now?	
23 (10111)	How <skip> you feeling now?	
25 (11001)	How are <skip><skip> now?	
27 (11011)	How are <skip> feeling now?	
29 (11101)	How are you <skip> now?	
31 (11111)	How are you feeling now?	

**iii. Comparison Among Various Detection Techniques**

among various detection techniques used to mark false identity described in Table II. Adikari and

In this section we discussed comparison analysis

**TABLE II: Comparison among various Detection Techniques**

Detection Technique	Advantages	Disadvantages
Winnowing + Orthogonal Sparse Bigram(OSB)	<ol style="list-style-type: none"> <li>Simple</li> <li>Easy to Implement</li> </ol>	Number of output nodes is required to be declared before start.
Neural Networks, SVM and PCA	Fairly good accuracy of 84% was achieved	Task could be performed on limited data only due to LinkedIn policies
OSB with SVM	Classifier could catch 99% Sybils	A limited number of accounts were found to be connected to other accounts. So there was less chance of any fake profile on Sybils(Renren).
OSB with check for regularity of tweets	Better accuracy since it uses all possible sequences for feature extraction	
Random Forest Classification based on various factors.	<ol style="list-style-type: none"> <li>Scalable</li> <li>Robust</li> <li>Experiment was performed on truly existing profiles</li> </ol>	

Dutta [6] have worked on identifying the fake profiles that are persisting on OSN like LinkedIn. This paper highlighted that by using technologies like Artificial Neural Network, Principal Component Analysis and Support Vector Machines, we can detect fake profiles with an accuracy of 84% which can be considered as a good detection considering a small and limited profile data as input. PCA, an acronym for Principal Component Analysis, is a linear dimension reduction tool. It is the 2nd order method based on covariance matrix of variables. The role of PCA is to find the linear orthogonal combinations, known as Principal Components in order to reduce the dimensions of data into linear dimensions.

Lee et al.[7] worked for the detection of fake profiles on OSN's MySpace and Twitter in 2010. They have collected the data of such 1500 on MySpace and 500 on Twitter. The parameters they have used is demographics, frequency of content generation, content, the number and type of connections. Stringhini et al.[8] in 2010 worked on the graph

based technique for fake profile detection instead of analysing all the profiles separately, analysis performed on Facebook and Twitter where 900 honeypot profiles are created and collected data from it for a period of over an year. Similarly C. Yang et al.[9] also worked on Graph based features in the year 2011 and also combined it with automation-based features and timing-based features too. Z. Yang et al.[10] used SVM classifier to detect fake accounts in Renren. The accuracy of the model they achieved is 99%. Cao et al.[11] and Conti et al.[12] have also worked on the detection of fake profiles using the graph features classification. Conti et al. [12] used the concept of analysis of distribution of number of friends. Cao et al.[11] connected their work to other fake profiles rather than connecting it to the real ones. Most of the people focussed on detecting the profiles which created spam or which can be classified as spam but in case of spamming, a group or mass is usually targeted. So it became a limitation that the methods used are not capable of detecting fake accounts involved in an Advanced Persistent Threat. This limitation was overcome by

Egele et al.[13] in 2015 who worked in the field of identifying the times when a good-high profile is encountered to act maliciously. He took into account, the behaviour of the account holders on OSN's, keeping a check on time and the place from where the messages were being received, the languages a person is connected to, his/her location, etc. He also targeted the similarities and dissimilarities between the user behaviour in order to classify them better. Then , they constructed a SVM classifier for it. This gave birth to a new trend of detecting fake profiles!

#### IV. Conclusion and Future Work

The detection of fake profiles through various researches have paved a way to adopt the best detection techniques. With reference to the researches , we are able to make comparisons among the mentioned techniques. By seeing the advantages and disadvantages for a number of techniques, we conclude that the technique of OSB and SBPH was found the best as it has the provision to extract a large number of features and a great accuracy score too. In future, we should try to build classifiers based on Unsupervised Learning and also take all the important things like friend request, language, comments,likes, etc. into consideration too for improvisation in results.

#### References

1. Andreas M. Kaplan, Michael Heinelein ESCP Europe, 79 Avenue de la Re ´publique, F-75011 Paris, France. *Two hearts in three-quarter time: How to waltz the social media/viral marketing dance* .
2. Kietzmann, J.H., Hermkens, K., McCarthy, I.P., Silvestre, B.S., 2011. *Social media? Get serious! Understanding the functional building blocks of social media*. *Bus. Horiz.*, SPECIAL ISSUE: SOCIAL MEDIA 54, 241– 251. doi:10.1016/j.bushor..2011.01.005.
3. Douceur, J.R., 2002. *The Sybil Attack*, in: *Revised Papers from the First International Workshop on Peer-to-Peer Systems, IPTPS '01*. Springer-Verlag, London, UK, UK, pp. 251–260.
4. Siefkes C., Assis F., Chhabra S., Yerazunis W.S. (2004) *Combining Winnow and Orthogonal Sparse Bigrams for Incremental Spam Filtering*. In: Boulicaut JF., Esposito F., Giannotti F., Pedreschi D. (eds) *Knowledge Discovery in Databases: PKDD 2004*. PKDD 2004. *Lecture Notes in Computer Science*, vol 3202. Springer, Berlin, Heidelberg
5. Chu, Z., Gianvecchio, S., Wang, H., Jajodia, S., 2010. *Who is Tweeting on Twitter: Human, Bot, or Cyborg?*, in: *Proceedings of the 26th Annual Computer Security Applications Conference, ACSAC '10*. ACM, New York, NY, USA, pp. 21–30. doi:10.1145/1920261.1920265.
6. Adikari, S., Dutta, K., 2014. *Identifying Fake Profiles in LinkedIn*, in: *PACIS 2014 Proceedings. Presented at the Pacific Asia Conference on Information Systems*.
7. Lee, K., Caverlee, J., Webb, S., 2010. *Uncovering Social Spammers: Social Honey Pots + Machine Learning*, in: *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '10*. ACM, New York, NY, USA, pp. 435–442. doi:10.1145/1835449.1835522.
8. M., Stringhini, G., Kruegel, C., Vigna, G., 2015. *Towards Detecting Compromised Accounts on Social Networks*. *IEEE Trans. Dependable Secure Comput.* PP, 1–1. doi:10.1109/TDSC.2015.2479616. Facebook, inc., 2014. *Facebook annual report* <https://www.sec.gov/Archives/edgar/data/1326801/000132680115000006/fb-12312014x10k.htm>.
9. Yang, C., Harkreader, R.C., Gu, G., 2011. *Die Free or Live Hard? Empirical Evaluation and New Design for Fighting Evolving Twitter Spammers*, in: *Proceedings of the 14th International Conference on Recent Advances in Intrusion Detection, RAID'11*. Springer- Verlag, Berlin, Heidelberg, pp. 318–337. doi:10.1007/978-3-642-23644-0\_17.
10. Yang, Z., Wilson, C., Wang, X., Gao, T., Zhao, B.Y., Dai, Y., 2011. *Uncovering Social Network Sybils in the Wild*, in: *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference, IMC '11*. ACM, New York, NY, USA, pp. 259–268. doi:10.1145/2068816.2068841.
11. Cao, Y., Li, W., Zhang, J., 2011. *Real-time traffic information collecting and monitoring system based on the internet of things*, in: *2011 6th International Conference on Pervasive Computing and Applications*. Presented at the 2011 6th International Conference on Pervasive Computing and Applications, pp. 45–49. doi:10.1109/ICPCA.2011.6106477.
12. Conti, M., Poovendran, R., Secchiero, M., 2012. *FakeBook: Detecting Fake Profiles in On-Line Social Networks*, in: *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012)*, ASONAM '12. IEEE Computer Society, Washington, DC, USA, pp. 1071– 1078. doi:10.1109/ASONAM.2012.185.
13. Egele, M., Stringhini, G., Kruegel, C., Vigna, G., 2015. *Towards Detecting Compromised Accounts on Social Networks*. *IEEE Trans. Dependable Secure Comput.* PP, 1–1. doi:10.1109/TDSC.2015.2479616.

# A Step towards protecting innocent civilian of internet terrene

Anvita Awasthi\*  
Harshit Sharma\*\*

## ABSTRACT

*The contribution of this research paper is an overview on cybercrime and the security issues related to this field. Centre of focus are the cyber issues that most people on internet are not aware of, cyber frauds in industries, effects of hacking on business, different methods used for cyber fraud in industries, latest cyber fraud issues and the prevention and security measures for people on internet and for industries as well.*

*This paper, reviewed on different backgrounds, will through light on those aspects, and attempts to make an awareness so people on internet can use the network wisely.*

### Thought Process

- Cyber – crime?
- Reasons for cyber attack
- Hactivity: hacking activity involved
- Deep web and cyber crime
- Security measures
- Conclusion.

### Cyber Crime

It is a term used mostly to describe criminal activity in which computers or computer networks are a tool, a target or a place of criminal activity.

A large portion of the Cyber crime issues are connected with the issues like security of budgetary dealings, giving security to data amid online exchanges, safeguarding protection and classification of messages and the assault on protection.

Survey report, 2014, of CERN [1]says 35% of the cyber-crime in organizations were not reported due to lack of evidence.

### Reasons for cyber attack

- Hackers
- Insiders
- Lack of knowledge
- Social engineering

Social engineeringThe practice of deceiving someone, either in person, over the phone, or using a computer, with the express intent of breaching some level of security either personal or professional.

### 419 fraud or Nigerian Scam

The 419 scam, better known as the Nigerian Scam, has grown into anepidemic. Basically an emailcomes to the target telling him he has been singled out for a very handsome deal and all he needs to do is offer a little bit of help. After the objective is sure and "signs on," an issue emerges that makes the objective pay a charge. After the expense is paid another issue comes up, alongside another charge. Every issue is "the last" with "one last charge" and this can be extended over numerous months. The casualty never observes any cash and loses from \$10,000–

\*(MCA 2nd year student), Institute of Management Studies (IMS), Ghaziabad., anvitaawasthi96@gmail.com

\*\* (MCA 2nd year student), Institute of Management Studies (IMS), Ghaziabad., harshit9430@gmail.com



\$50,000 USD all the while. What makes this scam so amazing is that in the past, official documents,

papers, letterhead, and even face-to face meetings have been reported.



### Hacktivity: hacking activity Involved

- Foot printing
- Phishing
- Social engineering
- DOS and D-DOS
- Sniffing
- Session Hijacking
- Hacking Webservers
- Hacking Web Applications
- SQL Injection
- Trojans and Backdoors
- Evading IDS, Firewalls, and Honey pots etc.

### Foot printing

- Foot printing, the initial phase in moral hacking, alludes to the way toward gathering data about

an objective system and its condition.

- Why footprint?
- Know security posture
- Reduce attack area
- Build database
- Draw network map.

### Phishing

**Wikipedia says:** Phishing is the attempt to obtain sensitive information such as usernames, passwords, and credit card details, often for malicious reasons, by disguising as a trustworthy entity in an electronic communication.





**Netcraft.**

Netcraft is an Internet administrations organization situated in Bath, England. Netcraft give web security administrations including hostile to misrepresentation and against phishing administrations. The Netcraft Extension is an apparatus permitting simple query of data identifying with the destinations you visit and giving security from Phishing.

**Key Features:**

- Protection against phishing sites
- Detailed site reports
- Risk Ratings
- Conveniently report suspected phishing & fraudulent site
- Protection against cross site scripting (XSS).



**Deep web and cyber crime**

Deep web: The Deep Web is any Internet content that, for different reasons, can't be or isn't recorded via web indexes like Google.

It is difficult to gauge because of covered up or bolted data inside databases. Early evaluates proposed that the profound web is 400 to 550 times bigger than the surface web.

- Malicious business exercises distinguished in the profound web, especially commercial centers and products cybercriminals trade.

- Deep web bolsters unlawful exchange of each kind of things.
- Drugs
- Stolen financial balances and so forth.

**Passports / Citizenship for sale**

There are several sites on the Deep Web claiming to sell passports and other forms of official ID, with prices varying from country to country, and seller to seller.

The screenshot shows the FakeID website interface. At the top, there is a navigation menu with links for 'Main', 'News', 'Services', 'Samples', 'log', 'Order', and 'Contacts'. Below the menu is a 'Pricing' section with a table listing prices for different countries. To the left of the table is an image of a passport. Below the table are images of passport covers for Canada, Finland, and Germany.

Country	Price for Passport	Price for Passport + Driving license	Price for Passport + ID card	Price for Passport + Driving license + ID card
Australia	600 Euro	700 Euro	700 Euro	800 Euro
Belgium	500 Euro	600 Euro	600 Euro	700 Euro
Brazil	400 Euro	-	-	-
Canada	600 Euro	700 Euro	700 Euro	800 Euro
Ireland	500 Euro	600 Euro	600 Euro	700 Euro
Italia	550 Euro	650 Euro	650 Euro	750 Euro

### Safety measures

Onion routing: The Internet is never truly anonymous. For the most part, your activities can be traced back to you; the only difference is how difficult it is to trace your actions back to you. That's where "onion routing" comes in.

Tor browser is a pre-configured web browser to protect your anonymity. Very much helpful while doing online transactions, net banking or sharing some sensitive information.

#### General security guidelines

- Use a full-service Internet security suite.
- Use strong passwords, don't repeat your passwords on different sites and make sure to change your passwords regularly.
- Keep all your software updated.
- Do not follow promotional links blindly.
- Manage your social media settings to keep most of your personal and private information locked down.
- Talk to your children and make sure they have some idea about any kind of online harassment, bullying or stalking.
- Do not connect to the public Wi-Fi.
- Most important: "be aware".

### Conclusion

- Cyber Crime is growing in wide scale and becoming a global issue.
- The Educational Institutes can assume essential part to make a solid moral base by including such subjects as obligatory ones.
- Government may do frequent checking on Cyber Community for illegal services and face them to strictly follow the standards.

Though there are some governmental and other organizations working to secure our digital world, we should also be aware of such situations and make aware of our natives aware so the internet terrene can use internet wisely and can stand strong together against cyber-crime.

### References:

- [1] *International Journal of Science and Research (IJSR)*, Volume 5 Issue 1, January 2016, *Cyber Crime: Critical View*
- [2] *Social engineering: Art of human hacking – Christopher Hadnagy*
- [3] *CEH – Introduction to ethical hacking.*
- [4] *Cybercrime in the Deep Web Black Hat EU, Amsterdam 2015*

# Security for SAP ERP System - A Case Study

Dr. C S Azad\*

## ABSTRACT

*Enterprise Resource Planning System (ERP) is the backbone of a number of large companies in the world. It has achieved an indispensable position in the daily operations of contemporary industry. SAP ERP is software of the German company SAP SE, a global market leader in ERP system. SAP ERP integrates the critical business functions of an organization. The ERP platforms allow companies to reduce costs, become more efficient, and respond faster to changes in the marketplace. Because of greater than before functionality in ERP, risk level has increased manifold. If not supervised appropriately, breach of security in ERP may cause momentous losses for a company. Security is critical for operation of the ERP system incorporating extremely confidential data of an organization. The present paper introduces the key security features of SAP ERP and presents the safeguards with the help of a case study.*

**Key Words:** SAP R/3, Authorization, Authentication, Role Based Access Control (RBAC), Segregation of Duties (SOD).

## Introduction

Enterprise Resource Planning System is one of the most important innovations in the world of Information Technology (IT). It has become a popular IT solution playing a vital role in the present business transactions.

Worldwide most of the organization uses the SAP ERP, the ERP software of SAP Germany.

SAP ERP embodies all business processes such as Operations (Sales & Distribution, Materials Management, Production Planning, Logistics Execution, and Quality Management), Financials (Financial Accounting), Supply Chain Management, Human Capital Management (Training, Payroll, e-Recruiting) and Corporate Services (Travel Management, Environment, Health and Safety, and Real-Estate Management).

ERP Systems consolidate the data at single location. This enhances the possibility of loss of valuable data in case of a security violation. There are significant issues about protection of database and system due to interconnection of business process. So a security concern in one department may lead to jeopardizing the entire ERP system.

The organization of this paper is as follows. First section describes the architecture of SAP ERP. Threats and security risks in SAP ERP are discussed in following section. This section also discusses the security aspects like Role Based Access Control (RBAC) and Segregation of Duties (SOD) in SAP ERP. Finally, security measures for SAP ERP system in BSNL are presented as case study. The scope of this paper is limited to application of appropriate security controls in SAP; the security of supporting infrastructure such as network, servers and PCs/Laptops is not covered here.

\*Asst. General Manager (ERP), BSNL, csazad68@gmail.com

## 1.0 SAP ERP Architecture

### 1.1 Enterprise Resource Planning (ERP) System

Enterprise Resource Planning (ERP) systems can be defined as business management systems that are module-based integrated software packages, controlling the smooth unified flow of information and processes across functional levels in an organization. The ERP system must contain financial and accounting module along with at least one other module with a shared database. (Elbardan and. Kholeif, 2017)

Kumar and Hillegersberg (2000) defined ERP systems as “information systems packages integrating information and information-based processes within and across functional areas in an organization”

German Corporation SAP SE is global market leader in the field of ERP. SAP R/3 (the "R" was for "Real-time data processing" and "3" was for "3 tier"): is the earlier name of the software developed by the German company SAP AG. In 2004, the latest version of the software, with revised technical architecture, was launched. This SAP ERP version is named as SAP ERP Central Component (ECC). The up to date version of the software is SAP ECC 6.0 Enhancement Pack 8.

#### 1.1.1 SAP ERP Architecture

SAPERP works in client server environment. Client – hardware/software environment that can ask for services for a central warehouse of resources. Server – hardware/software combination that can offer services to a group of clients in a controlled environment

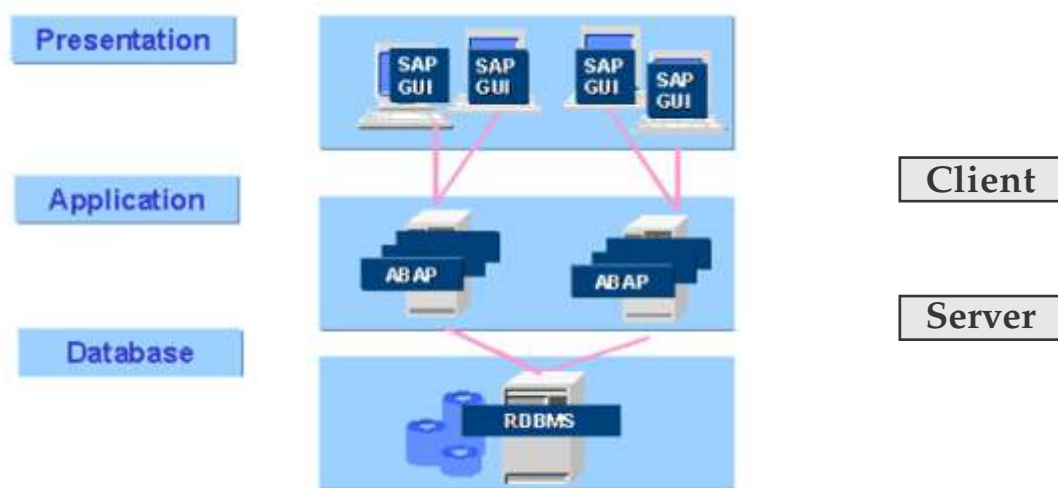


Fig. 1: SAP ERP Architecture  
(Reference: www.sap.com)

#### SAPERP Architecture consists of:

- i. Presentation layer on the client –side
- ii. Application server
- iii. Database server

#### i) Presentation Servers

Presentation servers contain systems competent to provide a graphical interface.

- Presentation Layer is also identified as client Layer
- Presentation Layer is a client interface

- In SAP-User interface is through Graphical user interface (GUI)
- Example – Desktop, laptops, Mobile Devices

#### ii) Application Servers

Application servers comprise dedicated systems with several Central processing Units (CPUs) and a huge quantity of Random Access Memory (RAM).

- The other names of Application Layer are Kernel Layer and Basic Layer.
- Application Layer is the layer where SAP application programs are carried out.

- The role of Application Layer is to communicate between Presentation and Database Layer.
- The work load is distributed to the different work processes in application server to accomplish the job.

**iii) Database Servers**

Database servers have dedicated systems with fast and large hard-drives.

- Data is stored in database layer.
- Example of data is Business data, SAP system data, SAP tables, Programs.
- Some of the data base servers are Oracle, Microsoft SQL Server and DB2 etc.

**2.0 Threats and Security Risk in ERP**

The mutually dependent character of ERP systems applications and the dependence on relational databases poses various business challenges to an organization that are reasonably different from conventional systems. (O’Leary 2000; Wright and Wright 2002).

The threats to the ERP system may be external and internal interrupters trying to access confidential information, amend data, create deceptive changes to programs, enter fraudulent transactions and perform other undesirable acts within the system. (Little & Best 2003).

**Little & Best (2003) have described the methods of unauthorized access in following ways:-**

- Passive Techniques** – To spy password, messages etc using electronic means.
- Attempted break ins** – Gaining access to system through users log in
- Masquerading** – When an intruder masquerades as an approved user using his login credentials
- Browsing** – When authorize user try to access sensitive data
- Viruses and worms** – Programs that invades the system and attempts to manipulate /destroy the data

Hunton and Writ (2001) established that the risk areas between ERP and non-ERP systems, differential risk assessments (ERP minus non-ERP) will be significantly higher in case of following:

- Business interruption risk;
- Network security risk;
- Database security risk;
- Application security risk;
- Process interdependency risk; and
- Overall internal control risk.

**2.1 Security in SAP ERP**

The basic information security principles are best summarized by the CIA triad: confidentiality, integrity and availability. Figure 2 demonstrates the CIA triad.



Figure 2: Information Security Triad

The following points are to be taken care of to make sure information security:-

- Confidentiality – Unauthorized revelation of information
- Integrity – Unauthorized change of data
- Availability – Disallowing service (a lack of accessibility to computing resources)

In SAP runtime environment, there is need to control both application security and unauthorized system access. Security is ensured in the SAP runtime environment by granting authorizations of roles to the user accounts. SAP authorizations manage access to Business Process Activities, or transactions within a particular business process.

The multiple layers of security in a SAP ERP system are as follows:

- Authentication – Only legitimate users should be allowed to log on the system.
- Authorization – Users should be allowed to carry out their designated transactions only. They should be restricted to specified places.
- Integrity – Protection of data from corruption or loss.
- Privacy – Security of data against illegal access.
- Obligation – Accountability and legal obligation towards stakeholders and shareholders together with substantiation.

The security in ERP system is based on the principle of Role Based Access Control (RBAC), Segregation of Duties (SOD), Basis Controls and Business Process Controls.

Role Based Access Control (RBAC):- All access is through roles. A role is basically a collection of permissions, and all users access the system only through the roles assigned to them.

Segregation of duties (SOD): -User roles are properly separated for carrying out business transactions. Segregation is based on the significance of information and transactions. User level and their specified duties are criterion for segregation of roles.  
Basis Control: - Basis control refers to technical controls related with user administration, password controls and direct access to data through tables.

Business Process Controls: - These controls refer to various automated and IT dependent controls

existing in SAP for a variety of business processes such as procurement, sales, HCM etc. These controls are inherent, configurable and procedural controls.

Each menu function in the SAP ERP system is assigned a transaction code. ERP Transaction codes are linked to authorization objects. For a user to be authorized to perform a T Code, he must have authorization in profile that is based on specific authorization objects.

An organization must segregate user duties properly to achieve internal control objectives. User authorizations can be used for segregation of duties in ERP systems. User profiles establish the access and authorization of the user and no user should be allowed to have inconsistent duties. Development, maintaining, and monitoring proper segregation of duties is responsibility of managers of the organisation. For this purpose a thorough study of individual job functions is required. The activities relevant to the duties of user must be identified properly. Monitoring user access in an ERP system will ensure that breaches of unauthorized access to the system are found and that procedures and employees with conflicting roles are quickly identified and those authorizations are terminated in an appropriate way. (Turner 2009)

### 3.0 Case Study: Security in BSNL ERP System

Bharat Sanchar Nigam Ltd. (BSNL), the public sector company implemented SAP ERP in 2009. With 50000 Dialog and 200000 communication users SAP licenses it was one of India's largest SAP ERP implementation project. A dialog user has access to SAP ECC while the access of communication users is limited to Employee Self Service portal. The ERP modules implemented in BSNL ERP are as follows:

- Finance & Control (FICO)
- Human Capital Management (HCM)
- Material Management (MM)
- Sales & Distribution (SD)
- Plant Maintenance (PM)
- Real Estate Management (REM)
- Project Systems (PS)
- Production Planning (PP)
- Strategic Enterprise Management (SEM)
- Quality Management (QM)
- Business Intelligence and Business Warehousing (BI & BW)



### 3.1 Basis Security Controls

- The user accounts that have been inactive for a period of 30 Days are disabled.
- Keeping record of all access violation and unauthorized logon attempts to the ERP systems. Violations, if any is reported and sternly dealt.
- All Information Systems logs are maintained securely and the access to these logs are provided only to authorize personnel.
- The access violation logs are reviewed on a regular (fortnightly) basis to take action to check their repetition. The user's access rights are reviewed on a monthly basis, to ensure that access has been revoked in case of transfer, resignation and terminations.
- Passwords are mandatory for all user accounts. The minimum length of password for user accounts is set as 8 alphanumeric characters. However, administrators or privileged users will have passwords at least 10 alphanumeric characters in length.
- User accounts are deactivated after three consecutive incorrect password attempts.
- The practice of "recycling" of reusing the same password when prompted for changed is prevented. A history of last used five passwords is retained for this purpose.
- Users are encouraged to change their passwords periodically.
- Time out session of inactive terminals is fixed for a specified time say fifteen minutes.
- A password expiration period of 30 days is set so that users are required to change the passwords on a monthly basis. However, the password for privileged accounts, like administrator accounts is changed once every fortnight.

### 3.2 Business Process Control

Inherent by default by SAP – Sale order cannot be created with an invalid customer similarly PO cannot be placed to invalid vendor.

Master Data Maintenance – Job of master data maintenance is assigned to designated teams. MDM teams at are responsible for integrity of important master data maintenance.

Configurable Controls – Critical business processes like procurement of goods / services (PO) and payment to vendors (MIRO) are subjected to approval through work flow. In PO approval finance controls checking budget availability, tax rate are mandatory while in MIRO engineer in charge is part of approval work flow. No vendor

payment using direct posting (FB60) is allowed.

Procedural Controls- IT dependent controls (For example – Review of Exception Tables)

### Conclusion

Any inappropriate access /unauthorized activity is biggest risk to the integrity of ERP systems. Whether they are innocent mistakes or fraudulent acts, it can seriously disrupt operations and incur financial loss. The security measures discussed in the case study can effectively check and protect the integrity of ERP system. These measures are worth emulating by the organizations for security of the SAP ERP systems.

### References:

- Elbardan H and. Kholeif AO (2017). Enterprise Resource Planning, Corporate Governance and Internal Auditing an Institutional Perspective, Springer International Publishing
- Genoulaz, V., Millet, P., & Grabot, B. (2005). A survey on the recent research literature on the ERP systems. *Computers in Industry*, 56, 510–522.
- Hunton, J., Wright, A. and Wright, S. (2001), "Business and audit risks associated with ERP systems: knowledge differences between information systems audit specialists and financial auditors", 4th European Conference on Accounting Information Systems (ECAIS), Athens
- Jones, M. C., & Young, R. (2006). ERP usage in practice: An empirical investigation. *Information Resources Management Journal*, 19(1), 23-42.
- Kumar K and van Hillebergersberg J(2000).ERP Experience and evolution . *Communication of the ACM*, vol.43 No.4
- Little, A., & Best, P. J. (2003). A framework for separation of duties in an SAP R/3 environment. *Managerial Auditing Journal*, 18(5), 419.
- Shehab, E. M., Sharp, M. W., Supramaniam, L., & Spedding, T. A. (2004). Enterprise resource planning: An integrative review. *Business Process Management Journal*, 10(4), 359-386.
- She, W., & Thuraisingham, B. (2007). Security for enterprise resource planning systems. *Information Systems Security*, 16(3), 152-163.
- Turner, L. D., & Owhoso, V. (2009). Use ERP internal control exception reports to monitor and improve controls. *Management Accounting Quarterly*, 10(3), 41-50.

# Movie Recommender System By Genetic Algorithm

Manoj Kumar\*  
Umesh Kumar Patel\*\*  
Mansi Gupta\*\*\*

## ABSTRACT

### Introduction

#### 1.1 Recommender Systems

Recommender systems or proposal systems (now and then supplanting "system" with an equivalent word, for example, stage or motor) are a subclass of data separating framework that look to anticipate the "rating" or "inclination" that a client would provide for a thing. Recommender systems are new sorts of web based programming apparatuses, intended to enable clients to discover their way through the present complex on-line shops and amusement sites. This portrays another recommender system, which utilizes a hereditary calculation to learn individual inclinations of clients and give custom-made proposals. This paper centers around the utilization of transformative hunt to tweak a profile-coordinating calculation inside a recommender system, fitting it to the inclinations of individual clients. This empowers there Recommender system to make more precise forecasts of clients' different preferences and thus better suggestions to clients..

#### 1.2 Aim and Objectives

The recommender system for the most part utilizes content based or collective or half and half systems for recommendation. The recommendations or ideal arrangement can be better gotten by genetic algorithms. Genetic Algorithms are stochastic pursuit strategies propelled from the instrument of normal development and genetic inheritance. Genetic algorithms chip away at a populace of

hopeful arrangements; every arrangement has a wellness esteem showing its closeness to the ideal arrangement of the issue. The arrangements having higher wellness esteems than others are chosen and furthermore get by to the people to come. Genetic algorithms at that point create better (i.e. new arrangements) by the blend of chose solutions. The strategies can find, safeguard, and engender promising sub arrangements. The proposed calculation would settle a two target, precision and assorted variety, streamlining issue of prescribing motion pictures to clients in view of their present inclination based genetic algorithm.:

- There are multiple objectives for recommender system - accuracy and diversity. Both these objectives are contradictory i.e. if one increases the other decreases and vice versa
- Data of the online movie recommendation platform (Movie lens) and open source language Python is used for building recommender system.

#### 1.3 Problem Definition

The recommender system portrayed is based around a community oriented separating approach, developing profiles of clients and after that utilizing a calculation to discover profiles like the present client. (Current client is utilized as dynamic client) Chosen information from those profiles are then used to fabricate proposals. Since profiles contain numerous qualities, a large number of which have inadequate or deficient information, the undertaking of finding suitable likenesses is regularly troublesome. To conquer these issues,

\*Asst. Professor, MIET, Meerut, mk.miet@gmail.com

\*\*Asst. Professor, IPEM, Ghaziabad, umesh.20778@gmail.com

\*\*\*Asst. Professor, IPEM, Ghaziabad, jv.mansi@gmail.com

current frameworks, (for example, Movie Lens) utilize stochastic and heuristic-based models to accelerate and enhance the nature of profile coordinating. This work makes such thoughts one stride further, by applying a developmental calculation to the issue of profile coordinating. In this examination, the Movie Lens dataset was utilized for starting tests. The developmental recommender framework utilizes 22 highlights from this informational index: motion picture rating, age, sex, occupation and 18 motion picture kind frequencies: activity, enterprise, liveliness, kids, comic drama, wrongdoing, narrative, show, dream, film-noir, frightfulness, melodic, secret, sentiment, spine chiller, and war, western.

## 2. Process

### 2.1 Recommendation Process

In general, each recommendation system takes after a particular procedure with a specific end goal to create item proposals, see figure 2.1. The recommendation methodologies can be arranged in light of the data sources they utilize. Three conceivable wellsprings of data can be recognized as contribution for the suggestion procedure. The accessible sources are the client information (demographics), the thing information (keywords, genres) and the client thing appraisals (obtained by transaction data, explicit ratings).

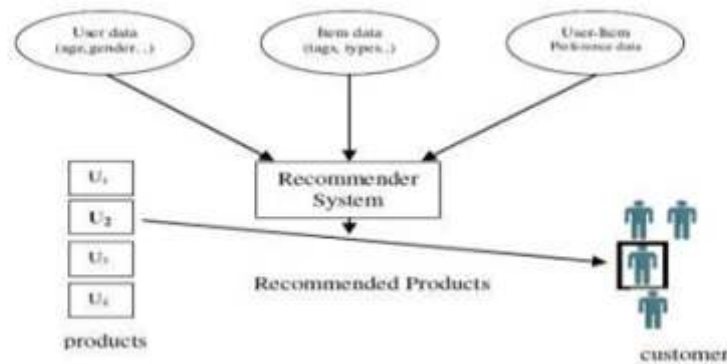


Figure 2.1. Recommendation process

### 2.2 Approaches of Recommender Systems

Recommender Systems (RS) are fundamentally coordinated towards people who need adequate individual experience or ability to assess the conceivably overpowering number of elective things that a Web website, for instance, may offer. One of the mainstream sites like Amazon.com utilizes a RS to customize the online store for every client. Since proposals are normally customized, distinctive clients or client bunches get different recommendations. Customized proposals are offered as positioned arrangements of things. In this computational assignment RSs attempt to anticipate what are the most appropriate items or administrations, in view of the client's inclinations and imperatives. Furthermore there are additionally non-customized proposals which are significantly less difficult to create and are ordinarily included in magazines or daily papers. Normal illustrations

incorporate the best ten determinations of books, melodies and so forth. Recommender systems utilized by different applications like E-commerce can be ordered in two ways:

- (i) Personalized recommender system
- (ii) Non-personalized recommender system

### 2.3 Functional Classification Of Recommender Systems

Recommender systems can be classified broadly into several categories depending on the information they use to recommend items. However the following four categories are mostly followed in the literature:

- Content-based Filtering Systems: In this system the filtering is done through used data about the items and information regarding the active user.

- Hybrid Recommender Systems: This method combine the prone and cons of two or more systems.
- Collaborative Filtering System: Make recommendations to the dynamic client utilizing data about an arrangement of clients and their connection with the thing. Demographic Filtering Systems: Use statistic data, for example, age, sex, instruction, and so forth of individuals for distinguishing sorts of clients.

- Does not require any subordinate data (which may not be accessible for some genuine issues).
- Is quicker and more proficient when contrasted with the conventional techniques.
- Have great parallel capacities.
- Optimizes both nonstop and discrete capacities and furthermore multi-target issues.
- Provides a rundown of "good" arrangements and not only a solitary arrangement.
- Always finds a solution to the issue, which shows signs of improvement over the time.

### 3. Genetic Algorithm

#### 3.1 Introduction

Genetic Algorithms are a group of computational models propelled by development. These calculations encode a potential answer for a particular issue on a basic chromosome-like information structure and apply recombination administrators to these structures as to save basic data. Hereditary calculations are frequently seen as capacity streamlining agent, despite the fact that the scopes of issues to which hereditary calculations have been connected are very wide. An implementation of genetic algorithm begins with a population of (typically random) chromosomes. One at that point assesses these structures and allotted regenerative open doors such that these chromosomes which speak to a superior answer for the objective issue are given more opportunities to 'reproduce' than those chromosomes which are poorer arrangements.

The 'goodness' of a solution is typically defined with respect to the current population. Genetic Algorithm (GA) is a pursuit construct enhancement system situated in light of the standards of Genetics and Natural Selection. It is often used to discover ideal or close ideal answers for troublesome issues which generally would take a lifetime to settle. It is much of the time used to take care of advancement issues, in look into, and in machine learning.

#### 3.2 Advantages of Genetic Allgorithm

Genetic Algorithm have various advantages which have made them immensely popular. These include

#### 3.3 Basic Terminology

Before beginning a discussion on Genetic Algorithms, it is essential to be familiar with some basic terminology:

- Population – It is a subset of all the conceivable (encoded) answers for the given issue. The populace for a GA is comparable to the populace for individuals with the exception of that rather than people, we have Candidate Solutions speaking to people.
- Chromosomes – A chromosome is one such answer for the given issue.
- Gene – A quality is one component position of a chromosome.
- Allele – It is the esteem a quality takes for a specific chromosome.
- Genotype – Genotype is the populace in the calculation space. In the calculation space, the arrangements are spoken to in a way which can be effectively comprehended and controlled using a figuring framework.

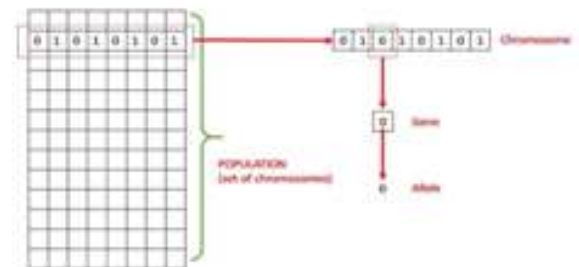


Figure 3.1: Genetic Algorithm Concept

- **Decoding and Encoding** – For basic issues, the phenotype and genotype spaces are the same. Regardless, in a vast part of the cases, the phenotype and genotype spaces are one of a kind. Unraveling is a methodology of changing an answer from the genotype to the phenotype space, while encoding is a technique of transforming from the phenotype to genotype space. Deciphering should be speedy as it is done on and on in a GA in the midst of the wellbeing regard estimation.
- **The Phenotype space** involves plans which essentially contain the thing amounts of the things to be picked. Regardless, in the genotype space it can be addressed as a twofold string of length  $n$  (where  $n$  is the amount of items). This is the place genotype and phenotype spaces are exceptional.
- **Fitness Function** – A wellness work essentially characterized is a capacity which takes the arrangement as information and produces the reasonableness of the arrangement as the yield.

Now and again, the wellness work and the target capacity might be the same, while in others it may be diverse in light of the issue.

- **Genetic Operators** – These change the hereditary creation of the posterity. These incorporate hybrid, transformation, choice and so forth.

### 3.4 Basic Structure

The basic structure of a GA is as follows –

We begin with an underlying populace (which might be created indiscriminately or seeded by different heuristics), select guardians from this populace for mating. Apply hybrid and change administrators on the guardians to produce new off-springs. Lastly these off-springs supplant the exist in people in the populace and the procedure repeat. One of the most important decisions to make while implementing a genetic algorithm is deciding the representation that we will use to represent our solutions.

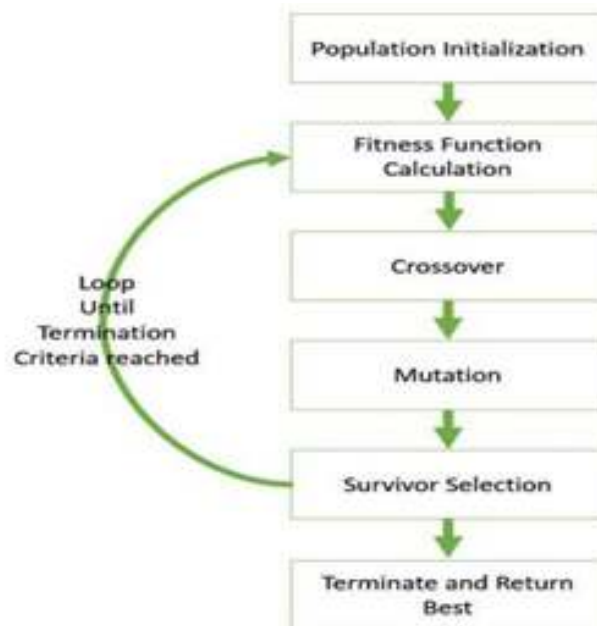


Figure 3.3: Basic process of Genetic Algorithm



4. Proposed Work

4.1 Problem Definition

The recommender system by and large uses content based or collaborative or cross breed methods for suggestion. The proposals or ideal arrangement can be better gotten by genetic algorithms. genetic algorithms are stochastic inquiry strategies enlivened from the instrument of normal development and hereditary legacy. genetic algorithms chip away at a populace of applicant arrangements; every arrangement has a wellness esteem demonstrating its closeness to the ideal arrangement of the issue. The arrangements having higher wellness esteems than others are chosen and furthermore make due to the people to come. Hereditary calculations at that point create better (i.e. new arrangements) by the blend of chose arrangements. The strategies can find, protect, and engender promising sub arrangements.

4.2 Proposed Work

The system described in this paper is based around a cooperative separating approach, developing profiles of clients and afterward utilizing a calculation to discover profiles like the present client. (Current client is alluded to as the dynamic client, A). Chosen information from those profiles are then used to assemble suggestions. Since profiles contain numerous qualities, a large number of which have inadequate or deficient information, the errand of finding proper likenesses is regularly

troublesome. To conquer these issues, current frameworks, (for example, Movie Lens) utilize stochastic and heuristic-based models to accelerate and enhance the nature of profile coordinating. This work makes such thoughts one stride further, by applying a trans formative (Genetic) calculation to the issue of profile coordinating.

In this paper , the Movie Lens dataset (<http://www.movielens.umn.edu>), is utilized for. The hereditary calculation based recommender system utilizes 22 highlights from this informational collection: motion picture rating, age, sexual orientation, occupation, and 18 motion picture classification frequencies: activity, experience, liveliness, kids, parody, wrongdoing, narrative, dramatization, dream, film-noir, ghashliness, melodic, secret, sentiment, science fiction, spine chiller, war, western. Open source dialect Python would be utilized for advancement of the recommender system.ml-100k dataset would be utilized from Movie Lens. This dataset has 100k evaluations from 943 clients on 1682 films.

4.3 Methodology

4.3.1 Generation of profile

Before recommendations can be made, the movie data is processed into separate profiles, one for each person, defining that person s movie preferences. We define profile(j,i) to mean the profile for user j on movie item i, see fig. 1. The profile of j, profile(j) is therefore a collection of profile(j,i) for all the items i that j has seen.

1. Rating    2. Age    3. Gender    4. Occupation ..... 22. Genre frequencies

5	23	0	45	000000100010000000
---	----	---	----	--------------------

Figure 4.1: profile(j,i) - profile for user j with rating on movie item i, if i has a rating of 5.

Once profiles are built, the process of recommendation can begin. Given an active user A, a set or neighborhood of profiles similar to profile(A) must be found.

Following files from ml-100k datasets are used - u.data, u.item and u.user. User id, age, gender and occupation come from u.user ; movie id and genre come from u.data and u.item has the user id, movie

id and rating values. Data from these 3 files is combined to create the profiles.

4.3.2 Finding the set or neighborhood of profiles

The success of a collaborative filtering system is very reliant upon the viability of the calculation in finding the set or neighborhood of profiles that are most like that of the dynamic client. It is crucial that,

for a specific neighborhood strategy, just the best or nearest profiles are picked and used to produce new suggestions for the client. There is little resilience for erroneous or superfluous forecasts. The area choice calculation comprises of three principle undertakings: profile determination, profile coordinating and best profile gathering..

**4.3.3 Nearest Profile Selection**

In a perfect world, the whole database of profiles would be utilized to choose the most ideal profiles. Be that as it may, this isn't generally a plausible choice, particularly when the dataset is extensive or if assets are not accessible. Subsequently, most frameworks pick irregular testing and this procedure is the duty of the profile choice piece of the algorithm.

**4.3.4 Figure out the closeness between the chose profile**

After profile determination, the profile coordinating procedure at that point figures the separation or closeness between the chose profiles and the

dynamic client's profile utilizing a separation work. Most present recommender frameworks utilize standard calculations that consider just "voting data" as the element on which the correlation between two profiles is made. However, all things considered, the manner by which two individuals are said to be comparable isn't construct exclusively in light of whether they have complimentary assessments on a particular subject, e.g., motion picture appraisals, yet additionally on different components, for example, their experience and individual points of interest. On the off chance that we apply this to the profile matcher and issues, for example, statistic and way of life data which incorporate client's age, sex and inclinations of motion picture classes should likewise be considered. Each client puts an alternate significance or need on each component. My approach indicates how weights characterizing client's needs can be advanced by a hereditary calculation.

A potential solution to the problem of evolving feature weights,  $w(A)$ , for the active user,  $A$  is represented as a set of weights as shown below in Figure 4. 2.

W1	W2	W3	.....	W22
----	----	----	-------	-----

Figure 4.2: Phenotype of an individual in the population

Where  $w_f$  is the weight associated with feature  $f$  whose genotype is a string of binary values. Each individual contains 22 genes, which are evolved by an elitist genetic algorithm (explained later).

The comparison between two profiles can now be conducted using a modified Euclidean distance function, which takes into account multiple features.  $Euclidean(A,j)$  is the similarity between active user  $A$  and user  $j$ :

$$euclidean(A, j) = \sqrt{\sum_{f=1}^{22} w_f * diff_f(A, j)^2}$$

Where:  $A$  is the active user

$j$  is a user provided by the profile selection process, where  $j \neq A$   $z$  is the number of common movies that users  $A$  and  $j$  have rated.  $w_f$  is the active user  $s$  weight for feature  $f$

$i$  is a common movie item, where  $profile(A,i)$  and  $profile(j,i)$  exists.  $diff_f(A,j)$  is the difference in profile value for feature  $f$  between users  $A$  and  $j$  on movie item  $i$ .

Note that before this calculation is made, the profile values are normalized to ensure they lie between 0 and 1. At the point when the weight for any component is zero, that element is disregarded. Thusly we empower highlight determination to be versatile to every client's inclinations. The distinction in profile esteems for occupation is either 0, if the two clients have a similar occupation or 1 generally.

ssss

### 4.3.5 Best Profile Selection

Once the Euclidean separations,  $euclidean(A,j)$ , have been found amongst  $profile(A)$  and  $profile(j)$  for all estimations of  $j$  picked by the profile determination process, the "best profile gathering" calculation is called. This positions each  $profile(j)$  as indicated by its similitude to  $profile(A)$ . The framework at that point just chooses the clients whose Euclidean separation is over a specific edge

esteem (considered most like the dynamic client) as the area of  $A$ . This esteem is a framework consistent that can be changed. To make a proposal, given a dynamic client  $A_n$  and an area set of comparable profiles to  $A_n$ , it is important to discover motion picture things seen (and preferred) by the clients in the area set that the dynamic client has not seen. These are then introduced to the dynamic client through a UI.

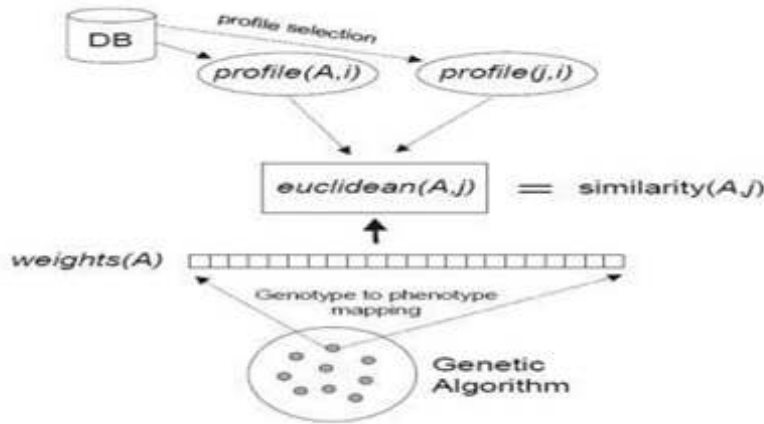


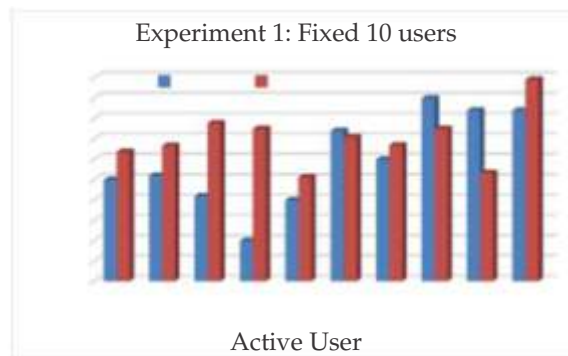
Figure 4.3: Calculating the similarity between A and j.

## 5. Result Analysis And Conclusion

### 5.1 Experiments

Four arrangements of analyses were intended to watch the distinction in execution between the transformative recommender framework and a standard, non-versatile recommender framework in view of the Pearson calculation. In each arrangement

of analyses, the anticipated votes of all the motion picture things in the test set (the things that the dynamic client has appraised yet were not utilized as a part of weights development) were registered utilizing the last element weights for that run. These votes were then looked at against those created from the basic Pearson calculation. The four arrangements of experiments were as per the following.:



Active User  
Figure 5.1

Experiment 1: Each of the first 10 users was picked as the active user in turn, and the first 10 users (fixed) were used to provide recommendations.

Graph above demonstrates the level of the quantity of appraisals that the framework anticipated effectively out of the aggregate number of accessible evaluations by the present dynamic client. While the expectations processed with the Pearson calculation dependably continue as before given a similar parameter esteems, those got from the GA shift per the component weights of that run. Out of the 10 keeps running for every dynamic client in each trial, the keep running with the best element weights (that gave the most astounding level of right expectations) was picked and plotted against the outcome from the Pearson calculation. In the main analysis, the GA recommender performed similarly well (or better) contrasted with the Pearson calculation on 7 dynamic clients out of 10. In the third test, out of the 50 clients the exactness for the GA recommender fell underneath that of the Pearson calculation for 17 dynamic clients. On whatever is left of the dynamic clients, the precision for the GA recommender was observed to be better – in some cases (user 16) the distinction was as incredible as 32%. The irregular testing for test 2 indicated awesome change on the forecast exactness for the GA recommender. Each of the 10 dynamic clients performed superior to the Pearson calculation.

## 5.2 Analysis of Results

Experiment 1 shows that the forecast exactness for the dynamic client 6, 8 and 9 on the GA recommender was more terrible than that acquired from utilizing the Pearson calculation. Be that as it may, when the quantity of clients was expanded to 50 in analyze 3, the exactness for the three specified dynamic clients climbed and beat the other calculation. This was normal – as the quantity of clients goes up, the likelihood of finding a superior coordinated profile ought to be higher and consequently exactness of the forecasts ought to likewise increment. The outcomes recommend that irregular inspecting is a decent decision for the profile determination errand of recovering profiles from the database. Arbitrary testing was relied upon to be superior to anything settling which clients to

choose on the grounds that it enabled the pursuit to consider a more prominent assortment of profiles (possibly 10\*10 runs = 100 clients in explore 2) and thus locate a superior arrangement of very much coordinated profiles. As specified before, just the run(s) with the best component weights for every dynamic client were considered for this examination. Taking a gander at the last component weights got for every dynamic client, numerous fascinating perceptions have been found Let's focus on a couple of active users – 4 and 27

Feature weights for active user 4

0 5 10 15 20 25

Figure 5.3

**The weights for feature 5-22 would be lower because of the scaling factor applied.**

Active user 4 is 24 year old male who is a technician by occupation. This user gives maximum preference to 2nd feature which is age. So it is likely that in this user's neighborhood other users with similar age group would be found. From the feature weights it can be seen that he gives more preference to War, thriller and horror movies which you would expect from a 24 year old boy. Another active user – 27 is analyzed who is a 40 year old female and is a librarian by profession.

This user gives more weight age to age and gender. She has interests in Western, sci-fi, romance, drama, crime and children's genres. She is a 40 year old female and so might have small children and that is she has interests in sci-fi and children's genres. She is a woman and so would like movies with romance and drama like most other women her age and given her profession.

## 5.3 Conclusion

This work has indicated how transformative inquiry can be utilized to adjust a profile-coordinating calculation inside a recommender system, fitting it to the inclinations of individual clients. This was accomplished by reformulating the issue of influencing proposals into an administered figuring out how to errand, empowering wellness scores to be registered by contrasting anticipated votes and genuine votes. Trials showed that, contrasted with a

non-versatile approach, the developmental recommender system could effectively tweak the profile coordinating calculation. This empowered the recommender system to make more exact forecasts, and subsequently better suggestions to clients.

### 5.4 Future Scope

I would like to use Archipelago based genetic algorithm where populations would evolve on different islands in parallel. This would aid in better exploration of feature weights. The algorithm would be applied on bigger datasets from movie Lens (the 1M and 10M datasets) to ascertain whether the algorithm scales. We can also try to introduce interactivity in the final solution – like presenting the user with initial movies based on his current mood and past/stored feature weights. The Genetic algorithm would then evolve the population over few generations before presenting the user with movie options. The user would select few movies and this would serve as synthetic fitness function for the Genetic algorithm. Within 2-3 such iterations the user would be presented the final list of recommended movies. We can also try to experiment with a different algorithm like Deep Learning based instead of Genetic Algorithm and compare the results.

### References

- [1] *Conducting Research Literature Reviews: From Internet to Paper*, by Arlene Fink, 2nd ed. Sage Publications: Thousand Oaks, 2005.
- [2] Hyun –tae kim, eungyeong kim, Jong hyun lee, Chnag wook Ahn, School of Information & communication Engineering, Sungkyunkwan University (SKKU), Cheonheon-dong, Suwon. 2010
- [3] Hybrid attribute-based recommender system for learning material using genetic algorithm and a multidimensional information model Mojtaba Salehi a,\* , Mohammad Pourzaferani b, Seyed Amir Razavi c, Received 6 May 2012; revised 6 December 2012; accepted 11 December 2012 Available online 3 January 2013
- [4] International Journal of Distributed and Parallel Systems (IJDPS) Vol.3, No.5, September 2012 DOI: 10.5121/ijdps.2012.3507 73 A CATEGORICAL REVIEW OF RECOMMENDER SYSTEMS RVVSV Prasad and V Valli Kumari Department of Computer Science & Engineering, Bhimavaram Institute of Engineering & Technology, Bhimavaram – 534243, India ramayanam.prasad@gmail.com .Department. Of Computer Science & Systems Engineering, AU College of Engineering, Andhra University, Vishakapatnam – 533003, India vallikumari@gmail.com
- [5] *Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions* Gediminas Adomavicius, Member, IEEE, and Alexander Tuzhilin, Member, IEEE IEEE transactions on knowledge and data engineering, vol. 17, no. 6, June 2005
- [6] Fournier, François. *Recommender Systems: Technical Report and Literature Review [Internet]. Version 19.* Knol. 2010 Oct 11
- [7] Ujjin, S., Bentley, P.J., *Building a Lifestyle Recommender System*, Poster Proceedings of the 10th International World, 2001
- [8] MacManus, R., *A guide to Recommender systems*, [http://www.readwriteweb.com/archives/recommendation\\_systems\\_where\\_we\\_need\\_to\\_go.php](http://www.readwriteweb.com/archives/recommendation_systems_where_we_need_to_go.php), 26 january 2009 A guide to Recommender Systems
- [9] Iskold, A., *Rethinking Recommendation Engines*, <http://alexiskold.wordpress.com/2008/02/25/rethinking-recommendation-engines/>, 25February 2008 Rethinking Recommendation Engines
- [10] Iskold, A., *The Art, Science and Business of Recommendation Engines*, [http://www.readwriteweb.com/archives/recommendation\\_engines.php](http://www.readwriteweb.com/archives/recommendation_engines.php), 16 January 2007
- [11] Ali K, van Stam W (2004) *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, TiVo: Making show recommendations using a distributed collaborative filtering architecture (ACM, New York), pp 394–401
- [12] Bennett J, Lanning S (2007) *Proceedings of the KDD Cup Workshop 2007, The Netflix prize* (ACM, New York), pp 3–6
- [13] X. Llorca et al, “Combating user fatigue in iGAs: partial ordering, support vector machines, and synthetic fitness,” in Proc. GECCO, pp. 1363–1370, 2005
- [14] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, “Evaluating collaborative filtering recommender systems,” *ACM Trans. Inf. Syst.*, vol. 22, no. 1, pp. 5–53, Jan. 2004.
- [15] Y. Koren, R. Bell, and C. Volinsky, “Matrix factorization techniques for recommender systems,” *Computer*, vol. 42, no. 8, pp. 30–37, 2009.
- [16] T. Zhou, Z. Kuscisik, J. G. Liu, M. Medo, J. R. Wakeling, and Y. C. Zhang, “Solving the apparent diversity-accuracy dilemma of recommender systems,” *Proc. Natl. Acad. Sci.*, vol. 107, no. 10, pp. 4511–4515, 2010.
- [17] Breese, J. S., Heckerman, D. and Kadie, C. 1998. *Empirical analysis of predictive algorithms for collaborative filtering*. In *Proc. of the 14th Conf. on Uncertainty in AI*, pp. 43–52.



# Image classification through Convolutional Neural Network

Anupama Sharma\*  
Shivam Dhama\*\*  
Vishal Rana\*\*\*

## ABSTRACT

We explored the problem of classifying images by the categorization of objects in different categories among a number of object categories. In today's world many machine learning algorithms are used in image classification one such classification method is deep neural network i.e. Convolutional Neural Networks (CNNs). CNN has accomplished best in class execution on an assortment of PC vision errands, especially visual order issues. In this paper, we build Convolutional Neural Network (CNN) integrated with feature extraction techniques to implement a general purpose 10 classes classification of images. We have used CIFAR10 dataset to conduct experiments. It contains 60000 pictures of 10 classes i.e. plane, vehicle, feline, puppy, winged animal, deer, truck, ship and steed in 6000 pictures for each class. To diminish overfitting in the completely associated layers of CNN, we utilized an as of late created regularization strategy called "dropout" that turned out to be exceptionally powerful. Our work is to train CNN on CIFAR10 dataset to classify 10 kinds of objects. In our experiments we have reached 80% of classification accuracy which is found significant in such classification problems.

## Introduction

In image classification problems, we investigate the numerical properties of different pictures and arrange information into categories [1]. Classification algorithms generally employ two phases of processing: a) Training and b) Testing Phase. In the training phase, characteristic properties of image features are extracted and based on these, a unique pattern is created. In the second phase, called testing phase, we predicted the classes of the images for the test set with the help of that correlation.

The two fundamental arrangement techniques are administered characterization and unsupervised order. In directed arrangement [2] each combine comprising of an information picture and a coveted class esteem. Whereas, in unsupervised classification there is no output value, we need to find some patterns and also identify some clusters so that we can predict the class of the new data (image).

There are lots of algorithms of machine learning and deep learning available to solve the classification problems. CNN is the a standout amongst the most famous algorithm. It is a profound learning calculation and it picked up fame through its utilization with picture information, and is as of now the cutting edge for recognizing what a picture is, or what is contained in the picture. CNNs even assume a basic part in undertakings like naturally producing inscriptions for pictures.

In this paper, we selected popular dataset CIFAR-10 that comprises of 60,000 pictures estimated 32 x 32 pixels. The dataset contains 10 classes that are totally unrelated (don't overlap) with each class containing 6,000 pictures. Finally, we used Convolution Neural Networks (CNNs) to identify the classes of CIFAR10 dataset.

Deep learning algorithms [3] require large dataset like CIFAR10 to learn more powerful models.

\*Asst. Prof. Ajay Kumar Garg Engineering College, Ghaziabad, anupama0027@gmail.com

\*\*Student, Ajay Kumar Garg Engineering College, Ghaziabad, shivamdharma95@gmail.com

\*\*\*Student, Ajay Kumar Garg Engineering College, Ghaziabad, vishalvedrana@gmail.com

During literature survey, we found Dropout is better technique for preventing overfitting hence we used Dropout in our experiments. We have achieved state-of-the-art performance in the classification task of the CIFAR-10 subset of the tiny images. We have used different combination of hidden layers and evaluate accuracy each time. Some combination give poor accuracy and some combination give very good accuracy. If dataset is small then any deep learning model gets overfit on training set hence may give good accuracy on train set but unable to generalize new observation, such type of models are called poor model. This paper implements a classification algorithm using easy-to-extract features and high efficient recognition algorithm. All features are extracted from tiny images; there are 3072 features per image.

The rest of this paper is organized as follows: section II introduces dataset, section III introduces to Convolutional Neural Network (CNN), section IV describes how CNN functions, Dropout is examined in segment V and trial comes about broke down in area VI talk about. At last section VII concludes paper with future directions of work.

**CIFAR10 Dataset**

CIFAR stands for Canadian Institute for Advanced Research. CIFAR10 is a set of natural color images of 32x32 pixels [17]. It contains 10 classes, each with 5000 training samples and 1000 test samples. Pictures fluctuate incredibly inside each class. They are not really focused, may contain just parts of the object, and show distinctive foundations. Subjects may shift in measure by a request of size (i.e., a few pictures indicate just the leader of a fledgling, others a whole feathered creature from a separation). Hues and surfaces of items/creatures likewise change extraordinarily. All through this paper we utilize the 60K modest pictures dataset, as  $32 \times 32$  shading pictures got via looking different online picture web indexes. Figure 1 demonstrates an arbitrary example of pictures from this dataset. For regulated adjusting, we have taken 6K pictures for each class. We considered 10 unique classes for our analyses. Those classes are airplane, automobile, bird, cat, deer, dog, frog, horse, ship, and truck. The classes are intended to be totally fundamentally unrelated for instance; neither vehicle nor truck contains pictures of pickup trucks.

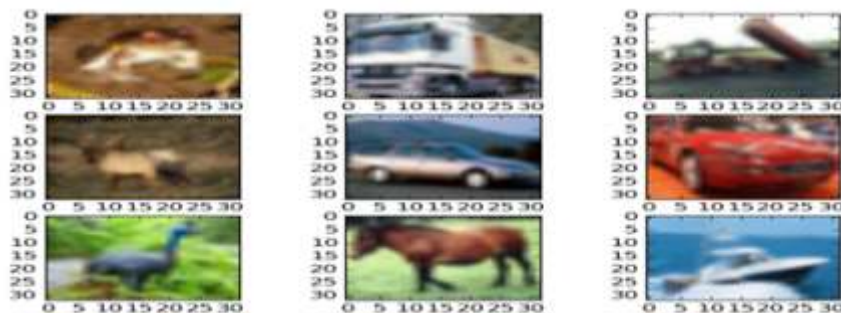


Figure 1: Image sample in CIFAR-10 DATASET

CIFAR-10 dataset consists of natural color images of size  $32 \times 32$  pixels in 10 classes of airplane, automobile, bird, cat, deer, dog, frog, horse, ship, and truck [23]. The training and test sets contain 50,000 and 10,000 images, respectively, and we hold out 5,000 training images as a validation dataset.

**3. Proposed work using Convolutional Neural Network (CNN)**

A neural system is a course of action of interconnected simulated "neurons" that exchange

messages between each other. The affiliations have numeric weights that are set in the midst of the planning methodology, so a genuinely arranged framework will respond successfully when given a photo or case to see. The system comprises of different layers of highlight recognizing "neurons". Each layer has various neurons that respond to different mixes of commitments from the past layers.

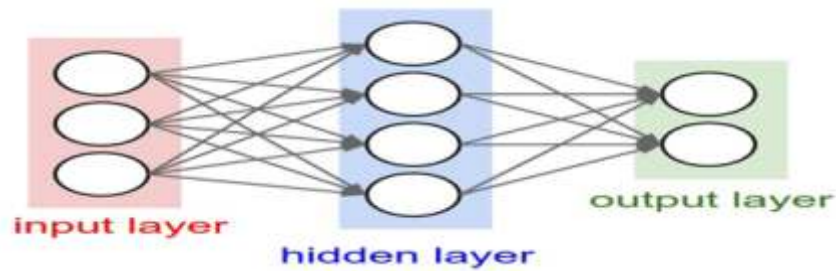


Figure 2: Neural Network

As shown in Figure 2, the layers are developed so that the principal layer recognizes an arrangement of crude examples in the information, the second layer identifies examples of examples, and the third layer distinguishes examples of those examples, et

cetera. Convolutional Neural Network is an exceptional sort of Neural Network that works similarly as a neural arrangement with the exception of that it has a convolution layer at the start.

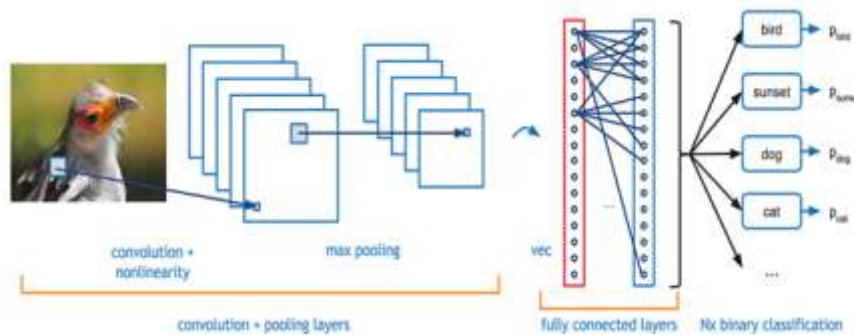


Figure 3: CNNs model

The image shown at left in figure 3, feeded as an input to the network, which goes through multiple convolutions, sub sampling, a fully connected layer and finally produced outputs. The image pixel is multiplied with feature detector (matrix of 3x3) which generates the feature map. The collection of feature map is called convolution layer. We created many feature map because we used different feature detectors. Each feature detector generated own feature map. The feature map contains only important information of the image. After that we have applied activation function [5] on each layer. We have used rectifier function (relu) for input and hidden layers and applied softmax function for output layer. After this we applied max pooling concept on feature map. The Pool size which we have used is (2x2) and then we applied flattening and make fully connected layer show in figure 3. The

final layer computes the output in the form of probability and gives the probability of each class respectively. Here we can say that the probability of  $P(\text{bird})$  should be higher than other class.

The main objective of sub sampling [6] is to get an input representation by reducing its dimensions, which helps in reducing overfitting. One of the strategies of sub examining is max pooling. With this method, you select the most elevated pixel esteem from a locale relying upon its size. After applying max pooling [7] we will get a single number that represents all the values in that window of the images. You use this layer to filtering: as the window moves over the image, you check for patterns in that section of the image. This works as a result of channels, which are increased by the qualities yielded by the convolution.

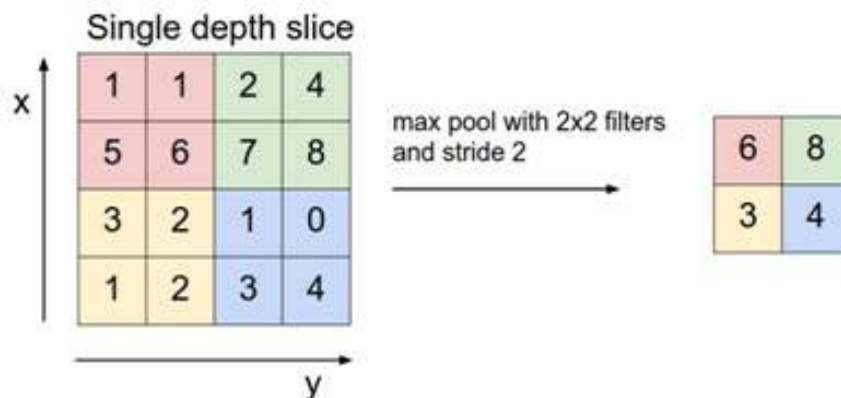


Figure 4: Max Pooling

The goal of the entirely linked layer is to flatten the high-level points that are learned through convolutional layers and combining all the features. It passes the flattened output to the output layer the place we use softmax classifier to predict the enter classification label.

#### 4. Dropout –Technique to reduce overfitting

In Neural networks if dataset is small then we generally suffer from overfitting problem. Due to

this model supply very true accuracy on coaching set and unable to generalize new observation. Dropout [8] prevents overfitting and provides a way of approximately combining exponentially many different neural network architectures efficiently. The time period “dropout” refers to dropping out devices (hidden and visible) in a neural network. By losing a unit out, we suggest briefly removing it from the network, alongside with all its incoming and outgoing connections, as shown in Figure 5.

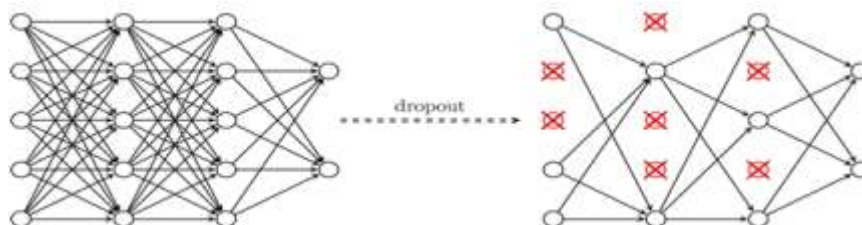


Figure 5:Left: A standard neural net with 2 hidden layers. Right: An example of a thinnednet produced by applying dropout to the network on the left. Crossed units have been dropped.

Applying dropout to a neural network results a “thinned” network from it. The thinned network consists of all the units that survived after dropout (Figure 5b).

#### 5. Experiment and Results

To each kind of class, 1000 pieces of images from testing sets are used to test the accuracy of our model. When we used a combination of hidden

layers [9] as 32, 64, 512, 10 (these are the no of nodes on each hidden layers separately), we got the accuracy of 70% that was not enough to classify the images very accurately. We tried the experiment with increased no of layers such as (32,32,64,64,512,1064,10) [10] and trained the model on 30 epochs and we observe an increased accuracy of 80% which is improved. Some species got a low accuracy. We can add more features to boost the accuracy of the model.

## 6. Conclusion

In term of classification, CNN models give the best result for CIFAR 10 image classification. Hence, we conclude that these algorithms are the most suitable for that particular task. On the other hand, the approach based on decision tree gives the worst results because of the overfitting problem. In general, a gaining knowledge of algorithm is stated to overfit relative to a less difficult one if it is more correct in becoming recognised data but less accurate in predicting new data. This paper introduces a neural network approach for image recognition. The computer can automatically classify 10 kinds of objects via the object images loaded from digital cameras or scanners. CNN is adopted for it has speedy pace on education and simple shape. Experimental result indicates that our algorithm is workable with accuracy greater than 80% on 10 kinds of objects of daily life. Future work is under consideration to improve the performance.

## References

- [1] Fuliang Wang And Feng Wang, "Rapidly Void Detection In TSVS With 2-D X-Ray Imaging And Artificial Neural Networks" *IEEE Transactions On Semiconductor Manufacturing*, Vol. 27, No. 2, May 2014.
- [2] *Supervised Machine Learning: A Review of Classification Techniques* S. B. Kotsiantis
- Department of Computer Science and Technology University of Peloponnese, Greece.*
- [3] Krizhevsky, I. Sutskever, and G. E. Hinton. *Imagenet classification with deep convolutional neural networks*. In *Advances in Neural Information Processing Systems 25*, pages 1106–1114, 2012.
- [4] Y. LeCun, K. Kavukcuoglu, and C. Farabet. *Convolutional networks and applications in vision*. In *Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on*, pages 253–256. IEEE, 2010.
- [5] A. Krizhevsky, 2009, *Learning multiple layers of features from tiny images*, Master's Thesis, Department of Computer Science, University of Toronto
- [6] A. Torralba and R. Fergus and W. T. Freeman, *80 Million Tiny Images: a Large Database for NonParametric Object and Scene Recognition*, IEEE PAMI, 2008
- [7] G. Hinton and R. Salakhutdinov. *Reducing the dimensionality of data with neural networks*. *Science*, 313(5786):504–507, 2006.
- [8] N. Srivastava. *Improving Neural Networks with Dropout*. Master's thesis, University of Toronto, January 2013.
- [9] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [10] Parekh, R., and Yang, J., and Honavar, V. (2000), *Constructive Neural Network Learning Algorithms for Pattern Classification*. *IEEE Transactions on Neural Networks*. 11(2), pp.436-451.



# BIG DATA & HADOOP: Characteristics and Importance

Gaurav Tyagi\*  
Gaurav\*\*

## ABSTRACT

Due to come of New Technologies, devices and lot of social media sites for communication, the large amount of data are used increasingly in every year. Big Data means really huge data, it is collection of large datasets that can't using traditional computing technologies.

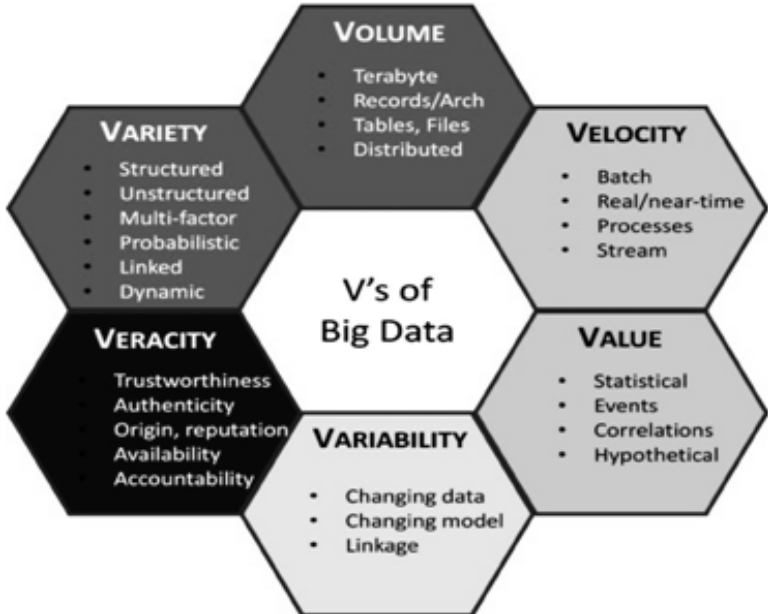
At the beginning of time means in 2003 data produce 3 billion gigabytes but in 2011 same amount of data created only in two days or in 2013 same data in every 10 min. Despite the fact that this data created is important and can be valuable when prepared, it is being dismissed. Now a days handling of big data is becoming a big issues for the users. This paper highlights the characteristics of big data and the importance and modules of hadoop.

**Keywords:** Hadoop, Modules, HDFS. Characteristics, Benefits

## Introduction

**Big Data:** Big data means there is a collection of data set which are very large and complex, which is

difficult to store and process during using available database management.



\*MCA student, IPED Gzb., gtyagi424@gmail.com

\*\*MCA student, IPED Gzb., stuartb683@gmail.com

**VOLUME:** It is defined the amount of data which use day by day at very fast process.

**VELOCITY:** It defined the pace of data in different source generate in every day or in every minutes.

**VARIETY:** It defined the various type of sources whose contribute the data which are generated in different area. It can be structured, semi- structured and also in un-structured.

**VALUE:** It defined that all of the data which are produced daily large amount of data is not useless it's all are valuable data which we are not ignore.

**VERACITY:** It characterized that the information in uncertainty or vulnerability of information accessible because of information irregularity and inadequacy.

### Hadoop

- It is an open source framework for running application on clusters of commodity hardware and store data from Apache and analyses the data whose are in large amount. It is written in JAVA and is not Online Analytical Processing (OLAP). It allows you to big data first in a distributed manner so you have to proceed parallely.
- Advancement of Hadoop started when ground breaking programming engineers understood that it was rapidly getting to be valuable for anyone to have the capacity to store and examine datasets far bigger than can for all intents and purposes be put away and got to on one physical stockpiling gadget (such as a hard disk).
- It was partially not fully because external storage become bigger. It take longer for the component that it read data from the disk to move a specific segment.
- About the majority of the enormous name of online locales name, and as everybody is allowed to refresh it for their own motivations, alterations made to the product by master engineers at, for instance, Amazon and Google or more related destinations which give bolstered back to the advancement group,

where they are frequently used to enhance the "official" item.

- This form of collaborative development between volunteer and commercial users is a key feature of open source software. At that time hadoop gets a lot of buzz in database. It is very helpful for the big data for distributed the data. But many people in the industry still don't really know what exactly Hadoop is.
- It is intended to scale up from a solitary server to a huge number of machines, with a high level of adaptation to non-critical failure. Instead of depending on top of the line equipment, the strength of these groups originates from the product's capacity to identify and handle disappointments at the application layer.

Why Hadoop is Important?

**Computing Power:** In the Hadoop's distributed computing model processing data are fast and there is no effect on the processing of the system means system are not going to be slow for produce big data.

**Fault Tolerance:** Data and application processing are protected against hardware failure. On the off chance that a hub goes down, employments are consequently diverted to different hubs to ensure the dispersed figuring does not come up short. Numerous duplicates of all information are put away naturally.

**Minimal effort:** The open source structure is free and uses product equipment to store vast measure of information.

**Versatility:** Anyone can undoubtedly develop your framework to deal with more information by including more hubs.

### Modules of Hadoop:

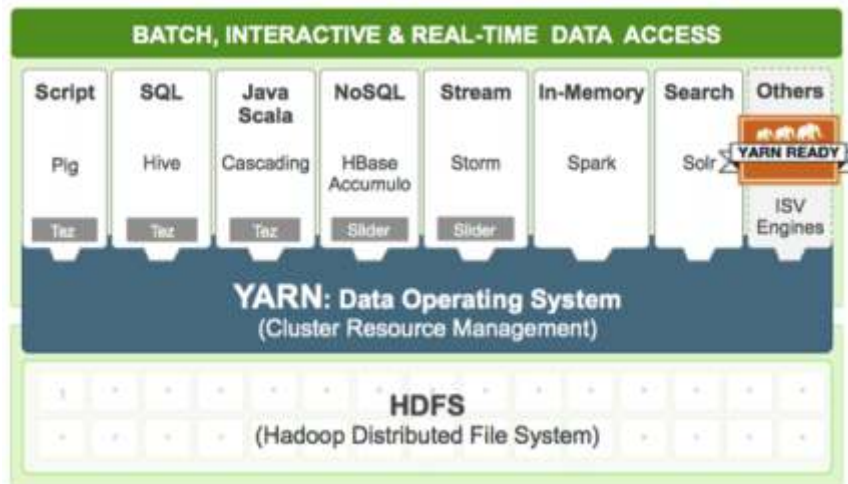
- HDFS (Hadoop Distributed File System)
- YARN
- Map Reduce
- Hadoop Common

1. **HDFS:** Hadoop Distributed File System is design to store large amount of data sets reliably and you can set at high bandwidth to user application.



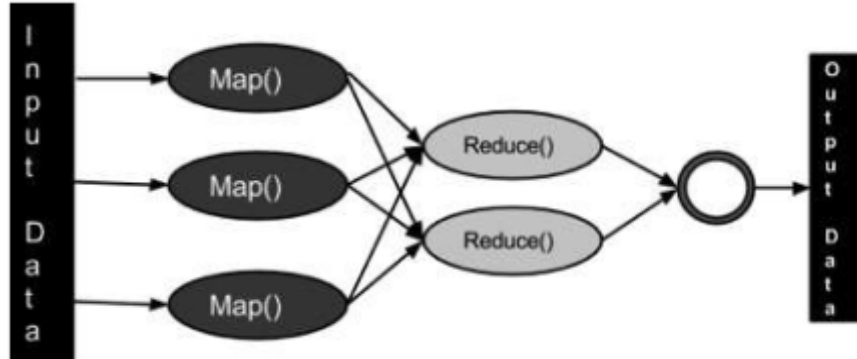
2. **YARN:** Yet Another Resource Negotiator is a cluster management technology and also used for job scheduling.
3. It is the essential for Enterprise Hadoop, giving asset administration and a focal stage to convey predictable tasks, security, and information

administration apparatuses crosswise over Hadoop groups. It's likewise stretch out the energy of hadoop to occupant and new innovations found inside the server farm with the goal that they can preferred standpoint of less cost and handling.



4. **Map Reduce :** It is software framework for easily writing application which process extremely big data and a program model for distributed computing based on java. Map take set of data which convert into another set of data, where elements are distributed into two parts (keys/value pairs). After this it reduces the task, which takes outputs from the map and

5. **Hadoop Comman:** Hadoop characterizes the accumulation of comparable/normal utilities and libraries that help other Hadoop modules. This bundle is considered as the base/center of the structure as it gives basic administrations and fundamental process.



### Conclusion

The accessibility of Big Data, ease item equipment, and new data administration and systematic programming have delivered a special crossroads in the historical backdrop of information examination. The meeting of these patterns implies that we have the capacities required to break down bewildering informational indexes rapidly and cost-adequately without precedent for history. These abilities are neither hypothetical nor insignificant. They speak to a certifiable jump forward and an unmistakable chance to acknowledge tremendous picks up as far as effectiveness, efficiency, income, and productivity.

The world is changing the way it is working presently and Big-information is assuming a vital part in it. Hadoop is a system that makes a

specialist's life simple while taking a shot at vast arrangements of information. There are enhancements on every one of the fronts. What's to come is energizing.

### Reference

1. [https://www.tutorialspoint.com/hadoop/hadoop\\_big\\_data\\_overview.htm](https://www.tutorialspoint.com/hadoop/hadoop_big_data_overview.htm)
2. <https://searchdatamanagement.techtarget.com/definition/Hadoop>
3. <https://www.quora.com/What-is-the-difference-between-big-data-and-Hadoop>
4. <https://www.techopedia.com/7/29680/technology-trends/what-is-the-difference-between-big-data-and-hadoop>
5. <https://www.javatpoint.com/what-is-big-data>
6. [https://www.sas.com/en\\_in/insights/big-data/hadoop.html](https://www.sas.com/en_in/insights/big-data/hadoop.html)

# INFORMATION SECURITY & CYBER CRIME: An inside story of Threats, Lacking & Solution.

Tushar Sharma\*

## ABSTRACT

*As we are living in the digitized world, our life is transforming into a digital/virtual framework. Our daily cycles FROM bed TO bed, everything we are doing in virtual world. Whether it is our education, job, business or any kind of banking transactions, everything is done on a virtual space known as cyber space. With this digitization, it is quite obvious that the threats will also rise exponentially. These threats are giving birth to a global concern known as cyber crime. The threat of virtual world has become real indeed. More the dependency more will be the threats & consecutively more will be the crime. This crime can be in any form like data theft, illegal access of data , intrusion into the devices, online frauds, etc.*

*This paper identifies the importance of being aware of the cyber crime effects, keeping mind the latest concerns that impacted the digital World & also provide the solutions to protect oneself from them. The present paper reviews the best way out to deal with the unforeseen situation of illegal cyber activities. This paper also include few case studies of cyber mishappening & some suggestions for future cyber space concerns.*

**Keywords:** Information Security, Cyber Space, Cyber Crime, Cyber Law ,Cyber safety, Threats, Hacking , intrusion, Digital world.

## Introduction

Cyber crime is related to the cyber space which includes the digital devices & the environment related to it. In generic sense, cyber crime is used which tends to cover all kind of civil &  $\neg$ [1] criminal wrongs related to computer.

Cyber crime can be classified under two headings:

- 1.1 Where computer is used as a tool to commit crime: The computer is a tool for an unlawful act where the offence reflects a modification of a conventional crime by making use of information technology & modern communication tools.
- 1.2 Where computer is the target of the crime: there are certain crimes where computer itself is the target.

There may be instances where the computer is a tool as well as the target of crime.

Another concern in computer crime is the possibility that an offender can commit multiple crimes in one go.

E.G. (A) In case of data theft, one has to hack the computer or any other electronic storage medium only the he can commit theft. The data theft includes hacking & theft.

(B) To initiate Distributed Denial of Service, installation of virus/Trojan horse on the slave system would be needed. The data of target system may also be altered or destroyed in the process. Thus DDoS includes hacking, introduction of virus & data alteration.[2]

\*Researcher, CO-FOUNDER & Speaker, TOFEE- "The Organization For Enlightenment & Education." New Delhi, India.,  
ddtusharsharma@gmail.com



Cyber crime includes data theft, illegal access of data, intrusion into the devices, online frauds, child pornography etc

**Discussion:**

The computer crimes can be categorized /classified into three categories:

1. Conventional Crimes through computer:- cyber defamation, digital forgery, cyber pornography, cyber stalking/harassment, Internet fraud, financial crimes, online gambling etc.
2. Crimes committed on computer network: hacking, denial of service.
3. Crimes relating to data alteration/destruction: virus/worms/Trojan horse/logic bomb, theft of Internet hours, steganography.

**Causes of Cyber Crime:**

Cyber criminals always choose the easiest way to get a big bounty. They generally targets big business personalities, business & organizations like banks, CA firms etc where the flow of funds is quite high. They try to fetch the sensitive data using various advanced techniques. Most of the times catching a criminal is quite difficult, the result of which the graph of cyber crime is gradually increasing. As computer systems are vulnerable, the safeguard mechanism is required to protect them against any kind of breach. Here are some of the reasons that we can list for the vulnerabilities:[3]

- 3.1. Easy accessibility: The complex technology has opened the doors for unauthorized access of computer systems. Sometimes it is very difficult to safeguard the privacy of a computer system(i.e. to prevent a computer system from unauthorized access).
- 3.2. Cyber Hoaxes: Cyber crime can be committed just to cause threats or damage one’s reputation. Among all the causes, this is one of the most dangerous one. They are called cyber terrorists[4].
- 3.3. Negligence: Negligence is a human tendency. So, most of the times, not paying proper attention invites cyber criminals to take access & control over the systems.
- 3.4. Tangled structure –Operating system is required to run a computer system & that operating system composed of millions of lines of codes. The human mind is imperfect, so

mistakes could occur at any stage. These loopholes help the cyber criminals in committing crimes.

- 3.5. Poor law Enforcing Bodies: Due to lack of effectiveness of cyber law, many Criminals get away without being punished. As committing crime is easy & getting caught is hard.
- 3.6. Revenge or Motivation: The lust to take revenge or to pull the lime light, most of the youngsters are driven towards illegal activities & the result of which they start tampering the data like fraud in bank transactions, booking airline tickets, e-commerce etc. They used to commit crimes for name, fame & publicity, without hurting anyone’s sentiment.[5]
- 3.7. Lack of Knowledge & Awareness: Lack of knowledge & awareness among most of the members of the investigating teams as well as the common man also helps the criminals in walking freely.
- 3.8. Bilateral agreements and international treaties: When the investigations lead to foreign countries, sometimes it sticks in between due to bilateral agreements & treaties, result of which the investigation ends with lack of evidences.

**Cyber Crime Investigation Report:**

The NCRB(National Crime Record Bureau) report on Cyber crimes in India for 2016 states all the details information about the type of offences, motives of crime & locations, where the crimes were committed. This report also states the acts or sections of the law they were tried under.

The report states the Cyber Crime status 19 Cities with more than 2 Million Population, 12317 cases were reported in 2016.[6]

Apart from the data, some Case studies have been included to elaborate on the threats and methods of defending against cyber attacks:

Case 1: Source Code Theft Case Indian IT company given bug fixing project for their source code by a US company One of the employees secretly took away the source code on two CDs Tried to sell the source code to other US companies. Demanded a price of US\$ 200,000

He received advance payment of US\$ 20,000 by wire transfer. A case was registered & the employee was

INFORMATION SECURITY & CYBER CRIME: An inside story of Threats, Lacking & Solution.

**Cyber Crime in Metropolitan Cities - 2014-2016**

S. No.	City	2014	2015	2016	% Share of City (2016)	Rank Based on Incidents/ % share (2016)	Actual Population (in Lakhs) (2011)	Crime Rate ++(Col.14/Population) (2011)	Rank Based on Crime Rate (2016)
1	2	3	4	5	6	7	8	9	10
1	Ahmedabad (Gujarat)	31	28	77	1.9	13	455	1.2	13
2	Bengaluru (Karnataka)	675	1042	792	58.1	2	82.0	8.0	1
3	Chennai (Tamil Nadu)	34	29	26	0.8	16	87.0	0.3	19
4	Coimbatore (Tamil Nadu)	20	20	18	0.4	18	21.5	0.8	16
5	Delhi	224	109	90	2.2	11	163.1	0.6	18
6	Ghaziabad (Uttar Pradesh)	61	25	62	5.5	14	22.6	2.6	10
7	Hyderabad (Telangana)	886	305	291	7.0	3	77.6	3.8	9
8	Indore (Madhya Pradesh)	63	65	23	0.3	17	21.7	1.0	15
9	Jaypur (Rajasthan)	374	363	532	12.8	3	30.7	17.3	2
10	Kanpur (Uttar Pradesh)	44	218	116	3.3	9	29.2	6.7	7
11	Kochi (Kerala)	63	26	22	0.8	15	21.2	1.6	11
12	Kolkata (West Bengal)	98	111	168	4.0	7	141.1	1.2	14
13	Kozhikode (Kerala)	11	10	16	0.4	19	20.3	0.9	17
14	Ludhiana (Punjab)	205	329	361	8.7	4	29.0	12.4	2
15	Mumbai (Maharashtra)	628	976	980	23.5	1	284.1	8.3	6
16	Nagpur (Maharashtra)	54	103	97	2.3	10	25.0	2.9	8
17	Patna (Bihar)	89	187	167	4.0	6	20.5	6.3	4
18	Pune (Maharashtra)	140	85	269	6.4	6	50.5	5.3	5
19	Surat (Gujarat)	79	84	66	1.6	13	45.8	1.4	12
<b>TOTAL CITIES</b>		<b>3265</b>	<b>4961</b>	<b>4172</b>	<b>390.8</b>		<b>1140.3</b>	<b>3.7</b>	

**Cyber Crimes - IPC Cases in Metropolitan Cities - 2016**

S. No.	City	Data Theft (Section 579 to 581)	Criminal Branch of Fraud/Fraud (Section 420, 424, 404)			Cheating (Section 420 IPC)	Forgery (Section 465, 467, 471A)	Other IPC Crimes	Total Cyber Crimes under IPC
			Computer Breach of Fraud/Fraud (Section 404, 405, 406) (Total)	Defaul Card/ Credit Card Fraud	Other Fraud				
1	Ahmedabad (Gujarat)	2	5	1	4	12	1	14	34
2	Bengaluru (Karnataka)	0	0	0	0	0	0	0	6
3	Chennai (Tamil Nadu)	0	0	0	0	0	0	0	0
4	Coimbatore (Tamil Nadu)	0	0	0	0	9	0	10	18
5	Delhi	0	0	0	0	10	2	12	24
6	Ghaziabad (Uttar Pradesh)	0	0	0	0	0	0	0	0
7	Hyderabad (Telangana)	0	0	0	0	0	0	0	0
8	Indore (Madhya Pradesh)	0	0	0	0	8	7	5	20
9	Jaypur (Rajasthan)	1	0	0	0	100	0	1	112
10	Kanpur (Uttar Pradesh)	0	0	0	0	22	0	0	22
11	Kochi (Kerala)	0	0	0	0	18	0	2	18
12	Kolkata (West Bengal)	0	0	0	0	0	0	0	0
13	Kozhikode (Kerala)	1	0	0	0	2	1	7	11
14	Ludhiana (Punjab)	0	0	0	0	0	0	0	0
15	Mumbai (Maharashtra)	6	15	7	8	711	22	196	906
16	Nagpur (Maharashtra)	2	0	0	0	48	0	20	70
17	Patna (Bihar)	0	0	0	0	0	0	0	0
18	Pune (Maharashtra)	1	1	0	1	152	0	26	180
19	Surat (Gujarat)	17	0	0	0	26	6	14	63
<b>TOTAL CITIES</b>		<b>36</b>	<b>21</b>	<b>8</b>	<b>13</b>	<b>1147</b>	<b>30</b>	<b>216</b>	<b>1513</b>

**Cyber Crimes - IT Act Cases in Metropolitan Cities - 2016**

S. No.	City	Tampering Computer Source Documents (Section 66)	IT Act - Computer Related Offences (Section 66 and Section 66B to 66E)					Cyber Terrorism (Section 66F)	
			Computer Related Offences (Section 66 and Section 66B to 66E) (Total)	Under Section 66	Under Section 66B	Under Section 66C	Under Section 66D		Under Section 66E
1	2	3	4	5	6	7	8	9	
1	Ahmedabad (Gujarat)	4	14	8	1	4	0	1	0
2	Bengaluru (Karnataka)	2	718	148	19	210	278	23	0
3	Chennai (Tamil Nadu)	0	23	6	0	7	3	4	0
4	Coimbatore (Tamil Nadu)	0	0	0	0	0	0	0	0
5	Delhi	1	46	22	0	26	8	2	0
6	Ghaziabad (Uttar Pradesh)	0	19	19	0	0	0	0	0
7	Hyderabad (Telangana)	1	284	82	0	0	202	0	0
8	Indore (Madhya Pradesh)	0	1	1	0	0	0	0	0
9	Jaypur (Rajasthan)	0	95	31	4	33	7	0	0
10	Kanpur (Uttar Pradesh)	0	94	67	0	27	0	0	0
11	Kochi (Kerala)	1	10	1	0	5	2	2	0
12	Kolkata (West Bengal)	3	18	3	0	35	21	0	0
13	Kozhikode (Kerala)	0	3	1	0	1	1	0	0
14	Ludhiana (Punjab)	13	346	180	0	131	7	0	0
15	Mumbai (Maharashtra)	0	24	13	0	3	8	0	0
16	Nagpur (Maharashtra)	0	15	1	0	9	5	0	0
17	Patna (Bihar)	0	187	187	0	0	0	0	0
18	Pune (Maharashtra)	2	17	18	1	17	21	0	1
19	Surat (Gujarat)	0	1	0	0	1	0	0	0
<b>TOTAL CITIES</b>		<b>29</b>	<b>2018</b>	<b>818</b>	<b>19</b>	<b>579</b>	<b>367</b>	<b>26</b>	

**Cyber Crimes - IT Act Cases in Metropolitan Cities - 2016 (Continued)**

S. No.	City	IT Act - Publication/Transmission of Offences / Security (Section 67, 67B, 67C, 67D)				Breach of Confidentiality/ Privacy (Section 72A)	Offence IT Act	Total Cyber Crimes under IT Act
		Publication/ Transmission of Offences / Security (Section 67, 67B, 67C, 67D) (Total)	Under Section 67B	Under Section 67C	Under Section 67D			
1	2	3	4	5	6	7	8	
1	Ahmedabad (Gujarat)	1	2	0	0	1	0	22
2	Bengaluru (Karnataka)	19	87	2	0	2	1	762
3	Chennai (Tamil Nadu)	3	9	0	0	0	0	26
4	Coimbatore (Tamil Nadu)	0	0	0	0	0	0	0
5	Delhi	11	14	1	0	0	1	94
6	Ghaziabad (Uttar Pradesh)	3	9	0	0	0	0	62
7	Hyderabad (Telangana)	4	4	0	0	0	2	250
8	Indore (Madhya Pradesh)	0	0	0	0	0	0	1
9	Jaypur (Rajasthan)	18	18	0	0	0	109	419
10	Kanpur (Uttar Pradesh)	20	20	0	0	0	0	114
11	Kochi (Kerala)	3	3	0	0	1	0	13
12	Kolkata (West Bengal)	24	24	0	0	0	0	177
13	Kozhikode (Kerala)	1	1	0	0	0	1	5
14	Ludhiana (Punjab)	0	0	0	0	0	0	361
15	Mumbai (Maharashtra)	0	0	0	0	0	0	24
16	Nagpur (Maharashtra)	2	2	0	0	0	0	18
17	Patna (Bihar)	0	0	0	0	0	0	167
18	Pune (Maharashtra)	19	18	0	1	0	0	88
19	Surat (Gujarat)	0	0	0	0	0	0	1
<b>TOTAL CITIES</b>		<b>140</b>	<b>154</b>	<b>2</b>	<b>1</b>	<b>4</b>	<b>390</b>	<b>2618</b>

S. No.	City	ILL Offences (Including Computer as Medium/Target)				Total Offences Under Cyber Crimes			
		Copy Right Act	Trade Marks Act, 1999	Other ILL Offences	Total Cyber Crimes under ILL	IT Act	IPC	SI	Total Cyber Crimes
		20	17	28	29	20	21	22	23
1	Amritsar (Punjab)	0	0	1	1	22	54	1	77
2	Bangalore (Karnataka)	0	0	0	0	762	0	0	762
3	Chennai (Tamil Nadu)	0	0	0	0	26	0	0	26
4	Coimbatore (Tamil Nadu)	0	0	0	0	0	18	0	18
5	Delhi	0	0	0	0	66	24	0	90
6	Ghaziabad (Uttar Pradesh)	0	0	0	0	62	0	0	62
7	Hydrabad (Telangana)	0	0	0	0	291	0	0	291
8	Indore (Madhya Pradesh)	0	0	0	0	1	20	0	21
9	Jipur (Rajasthan)	1	0	0	1	419	112	1	532
10	Kanpur (Uttar Pradesh)	0	0	0	0	114	22	0	136
11	Kochi (Kerala)	0	0	0	0	15	18	0	33
12	Kolkata (West Bengal)	0	0	0	0	388	0	0	388
13	Kochi/Kode (Kerala)	0	0	0	0	9	11	0	20
14	Lucknow (Uttar Pradesh)	0	0	0	0	361	0	0	361
15	Mumbai (Maharashtra)	5	1	0	6	24	890	6	921
16	Nagpur (Maharashtra)	0	0	0	0	18	79	0	97
17	Patna (Bihar)	0	0	0	0	367	0	0	367
18	Pune (Maharashtra)	0	0	1	1	88	180	1	269
19	Surat (Gujarat)	0	0	0	0	1	65	0	66
<b>TOTAL CITIES</b>		<b>6</b>	<b>1</b>	<b>1</b>	<b>8</b>	<b>2410</b>	<b>1332</b>	<b>8</b>	<b>4150</b>

S. No.	City	Illegal Sale	Revenge	Insult to Modesty of Women	Sexual Exploitation	Extortion/Black-mailing	Fraud / Satisfaction of Banning Carded	Coercing Sharepays	Inviting Hate Crimes Against Community
1	2	3	4	5	6	7	8	9	10
1	Amritsar (Punjab)	48	0	0	8	1	1	0	0
2	Bangalore (Karnataka)	584	27	24	91	27	2	6	0
3	Chennai (Tamil Nadu)	10	1	1	2	0	0	4	1
4	Coimbatore (Tamil Nadu)	8	3	0	0	0	0	0	7
5	Delhi	34	0	10	1	0	2	1	0
6	Ghaziabad (Uttar Pradesh)	18	13	0	12	1	1	0	0
7	Hydrabad (Telangana)	74	217	0	0	0	0	0	0
8	Indore (Madhya Pradesh)	13	0	2	1	0	0	1	0
9	Jipur (Rajasthan)	190	45	0	1	22	0	4	0
10	Kanpur (Uttar Pradesh)	78	0	0	10	14	0	0	19
11	Kochi (Kerala)	20	9	1	1	0	0	7	0
12	Kolkata (West Bengal)	34	35	20	14	30	0	0	0
13	Kochi/Kode (Kerala)	12	0	2	1	0	0	0	0
14	Lucknow (Uttar Pradesh)	270	0	0	0	0	91	0	0
15	Mumbai (Maharashtra)	688	5	207	89	12	0	0	1
16	Nagpur (Maharashtra)	21	2	4	0	0	0	1	0
17	Patna (Bihar)	107	40	0	0	0	0	0	0
18	Pune (Maharashtra)	110	24	0	18	9	1	12	1
19	Surat (Gujarat)	61	0	0	0	0	0	1	0
<b>TOTAL CITIES</b>		<b>2280</b>	<b>404</b>	<b>234</b>	<b>140</b>	<b>136</b>	<b>302</b>	<b>14</b>	<b>20</b>

Police Disposal of Cyber Crime Cases in Metropolitan Cities – 2016

S. No	City	Cases Pending Investigation from Previous Year	Cases Reported During the Year	Total Cases for Investigation (Col. 3+4)	No of cases withdrawn by the Govt	Cases transferrred to other Police Station or Magistrate	Cases Not Investigated U/S 157(1) (b) of Cr.PC	Final Report		
								True but insufficient Evidence	False	Mistake of Fact
1	2	3	4	5	6	7	8	9	10	11
<b>TOTAL CITIES</b>		<b>4687</b>	<b>4172</b>	<b>8859</b>	<b>0</b>	<b>22</b>	<b>0</b>	<b>1749</b>	<b>73</b>	<b>340</b>

Police Disposal of Cyber Crime Cases in Metropolitan Cities – 2016 (Concluded)

S. No	City	Final Report		Cases in Which Charge-sheets were Submitted	Total Cases Disposed off by Police (Col. 7+8+ Col. 13+ Col.14)	Cases Pending Investigation at the End of the Year (col. 5 - (Col.6 + Col.15))	Charge-Sheeting Rate (Col.14/ Col.15) * 100	Pendency Percentage (Col.16/ (Col.5)) * 100
		Non Cognizable	Total (Col.9+ Col.10+ Col.11+ Col.12)					
12	13	14	15	16	17	18		
<b>TOTAL CITIES</b>		<b>12</b>	<b>2174</b>	<b>813</b>	<b>3009</b>	<b>5850</b>	<b>27.0</b>	<b>66.0</b>

arrested in a Delhi 5 star hotel. Case charge-sheeted, presently pending trial.

Case 2: In July 2015, an Azerbaijani woman was blackmailed by an India-based man. She met him on Facebook and, at his convincing, sent him her pictures. The man then threatened to morph the pictures and upload them to porn websites if she did not send him \$100 every month. Apparently to show he wasn't bluffing, the man uploaded a few of the woman's morphed pictures to Facebook, forcing her to deactivate her account.

Case 3: Job Fraud[7] A Email Received by the Victim which posed to be from Maruti Suzuki (info@marutisuzuki.com) (SPOOFED) that his resume has been shortlisted from a Job Site Monster.com for engineer at MARUTI SUZUKI PLANT offering him a salary of Rs.2.0 lacs /month.He has to deposit Rs.8,200 in a STATE BANK OF INDIA Account Number and come for the interview with the pay slip and also that it was said in the email that this amount is refundable.The Emails traced were from all foreign countries , and the BANK Account were also fake to which the money was deposited and the amount was immediately withdrawal from ATMs

As per the report, among 29 states in India, Uttar Pradesh was on the top with Maharashtra following to the second position in cyber crime. The two states were on same level in 2015, and Maharashtra was one step up than Uttar Pradesh in 2014.[8]

Most of the cases are still pending due to insufficient evidences. In most of the states even the teams taking care of these type of cases are not well trained to deal with these matters. The cases registered in the police stations are either forwarded to Cyber cells or other dedicated teams of that state or just kept as a diary entry in the stations.

Person who is exploited, looks for the solution only. He/she does not want to go for the procedure of complaint. He /she look for the solution only, without even realizing that they are the victims of a criminal acts.

## Defending Against Cyber Crimes

As per the above case studies & the report published by NCRB in 2017, there is a need to be more attentive & more careful while online or using digital medium. As far as law enforcement agencies are concerned, prevention of crime is more important & one of the priority than the detection of one after it has occurred.

If internet is accepted as a medium of communication and publication and exchange of ideas, the concern here must be that any form of preventive measures should be minimal. Otherwise ,prevention method may run into difficulties of "Voilation of Privacy". Prevention of online crime definitely needs a different & hi-tech approach than the real world, some of these approaches are:

### Technology as Aid to Prevention.

High Technology crime must be prevented using high technology. Larry Lessig contended that cyber space can be regulated through law & programming code. In cyberspace,internet browsers can be configured to prevent repeated password entry attempts for sensible data.[9]

Encryption is the another way by which crime can be prevented.[10]

### User Awareness

Since computers are in the possession of the victims, awaring them about the security measures can prove to be one of the best means of preventing crime on internet.

The greatest security threat to computer systems is from insiders. Studies reveal that over 80% of the computer theft is committed from within the organizations. Keeping a check on one's own employees can also help in preventing such offences. The IT Act & Prevention of Offences

The IT Act has also conferred power on the police to prevent the commission of offences under the Act. Section 80(1).[11]



## Conclusion

In this Virtual world as we are becoming more slave to the technology, our dependency is also increasing day by day & a result of that the challenges to be secure in the virtual world is also a biggest challenge for everyone. This paper is an approach to highlight some crucial aspects of cyber crime that are knowingly/unknowingly ignored by us. New cyber laws & policies must be implemented to tackle with cyber crime but knowledge and awareness of existing cyber policies & laws is also required. A famous proverb "Prevention is better than cure" Need to be understood by individual as well as the organizations to avoid any unforeseen situation.

## References:

- [1] David R. Johnson & David Post. "Law & Borders- The Rise of Law in Cyberspace". *Stan L Rev* 48(1996):1367,1372-75.
- [2] Kshetri, Nir (2005) "Pattern of global cyber war and crime: A conceptual framework, " *Journal of International Management, Elsevier, vol. 11(4), pages541-562*
- [3]. <https://krazytech.com/technical-papers/cyber-crime>
- [4] Michael Massourakis & Farahmand Rezvani & Tadashi Yamada (1984) "Occupation, Race, Unemployment and Crime In a Dynamic System, " *NBER Working Papers 1256, National Bureau of Economic Research, Inc.*
- [5] Panu Poutvaara & Mikael Priks (2005) "Violent Groups and Police Tactics: Should Tear Gas Make Crime Preventers Cry?, " *CESifo Working Paper Series 1639, CESifo Group Munich.*
- [6] NCRB Report 2017 <http://ncrb.gov.in/StatPublications/CII/CII2016/pdfs/NEWPDFs/Crime%20in%20India%20-%202016%20Complete%20PDF%20291117.pdf>
- [7]. <http://gurgaon.haryanapolice.gov.in/job-froud.htm>
- [8]. Tom Standage. *The Victorian Internet*(1998):100-107.
- [9]. 6 No. 6 *Cyber Crime Law Reporter* 6 <http://www.usdoj.gov/usao/fls/060210-03.html>
- [10] Lawrence Lesser. "Code and Other Laws of Cyberspace.:(1999):53-60 quoted in *ibid.*
- [11] *Information Technology Act 2000*
- [12] <https://scroll.in/article/860254/what-the-national-crime-records-bureau-report-does-not-tell-us-about-cyber-crime-in-india>
- [13] <http://www.firstpost.com/tech/news-analysis/ncrb-report-shows-trends-in-cybercrime-in-cities-states-and-union-territories-and-contains-some-surprising-insights-4242851.html>



# Use of Augmented Reality to build Various Application making world more interactive and informative

Priyank Mathur\*  
Harshit Kumar\*\*

## ABSTRACT

*The aim of this project is to make world more interactive and informative using augmented reality. Augmented reality is that could make our complex life easy such as to ease implementation of complex operation. Augmented Reality mixes virtual environment with actual reality, making knowledge available to the user ; tools to ensure efficiency in the transferring knowledge for various processes in different environment.*

*This technology will truly change the way in which we view the world. Augmented reality adds some sort of sound graphics, and fragrance to the real world as it exists.*

*The Benefit of this technology is, its ability to place computer generated graphics in the field of user's vision. With the help of advanced augmented reality technology, for example, 3d object viewer, dynamic Visiting card etc.*

*General Terms : Real life, Mixed reality, Head tracking, Field of view , Aspect ratio etc.*

*Keywords : Image Rendering , Markers , tagging , AR overlays*

## Introduction

Augmented Reality has evolved from a head wear device to a cell phones. It presents implementation on the small screen device has shaped AR to what it is now a days. There are some challenges since this the technology which directly dependent on the growth of digital era and computational technology. It totally undefined in term of potential how far it will go in future but will bring good to humans .Software , Hardware and a Remote server are the three basic components of the AR technology. Software plays a key role in the efficient functioning of any Augmented reality device. Some Special Augmented reality programs are used in AR-based applications like D'Fusion. Virtual images are overlapped with the real live image, and that can be generated using some sort of 3D software. Software like AutoCad3D, Studio Max or Cinema4D..To experience Augmented reality, end-user is required

to download the software application for that specific operating system.

## Problem Statement

Our view of the world should not be limited by what our simple eye can see. There are Gadgets that we use to gather and view our related information over it, real time access to data and most importantly, a storage mechanism to store it as memories. Why not play or display the data or digital media over physical world. Yes, definitely it's possible, and Augmented reality is here to stay. The potential for this technology is undoubtedly very high, that has mostly ranged from almost any field . Nothing ground breaking till now – no mass adoption of such wearable technology or any of related implementations have become part of our day to day lives.

\*Computer Science, IMS Engineering College, priyankmathurbtech@gmail.com

\*\*Computer Science, IMS Engineering College, deecoderr@gmail.com

Augmented Reality, just like other technologies will have some crucial uses and will fruit full in coming future. However, all forms of technology will have come risks associated with them . There are several platforms to create application for augmented reality like Ar Tool kit , Unity ,Vuforia sdk etc in market. Using these technology we create application based on idea that would make day to day life easy and interactive by solving real life problems.

### Significance of Problem in Real World & Applications

1. **Profiling:** Integrating various technology along with augmented data that lead to a vast integration of application in our lives].
2. **Unauthorized Advertising (Augmented):** Several Advertisers and other companies are trying the possibilities of monetizing objects that exist in the actual world .This is done by placing some digital advertisements on to physical real-time.
3. **Augmented Behavioural Targeting:** Their are several algorithm to make image target locate and decide what to do with the surrounding . With various conditions of changing surrounding to make the relevant data augment is still a problem.
4. **Physical danger:** With new technology comes various hidden dangers . With this , their are devices to wear or use. There are chances of short circuit , eyes also get strain with continuous viewing as well.
5. **Spam:** where there is a business opportunity, there are chances of spam, deceptive advertisements and software build tricks to fool consumers paying for things they don't need.

### Objective

- Planting digital knowledge to real-world scenario
- Closer to the actual world.
- Adds bit of information and meaning to the physical object.
- Physical world is made more useful by digital matter
- Blended experience.
- Playback videos on the visiting card of the user .

- Recognizing text written anywhere in the live scene.
- Elaborate information about anything by augmenting the extra information on the real world scene.

### Related Work

In this section, we will briefly review about previous work done to check whether the pre-existing design models and methodologies address the needs of AR interfaces. And what needed to be done to make advanced modules.

A known user interface design model which is the Four Level approach that is developed by Foley and colleagues. As described by this model the user interface acts as medium which provides a dialogue in between the user and the computer. These four levels are arranged based on the meaning and the form of the dialogue in between the user and the computer. These levels mostly focuses on user interactions through explicit commands. And this approach works fine for command language and GUI style interfaces. But is not enough to meet AR interface requirements like implicit interactions, physical objects. and object dynamism.

Another related user interface design model called Command Language Grammar (CLG) that was developed by Moran[24].It provides platform for designers a model to design and describe command language interfaces. As in Foley's model, CLG has divided the interface design into levels. But specifications in the levels are more formal and detailed in CLG. But CLG works good in command language interfaces, its applicability to AR interfaces is limited., interactions using implicit commands, Object dynamism , physical objects and commu0nication patterns among these objects are out of the scope with this model.

A third interface design model called Shneiderman's Object Action Interface model . It is developed specifically for design of GUI style interfaces. To meet the needs of GUI style interfaces, this model emphasizes on the importance of visual representations of objects and their respective actions.This

model basically focuses on explicit command style interactions using direct manipulations, keeps the amount of syntax small in interaction specifications samples. However, OAI doesn't address the distinctive characteristics of AR interfaces such as object dynamism, physical objects ,implicit style interactions, and communication patterns among objects.

Another technique for designers is to implement general-purpose design models and methodologies like object oriented design model and Object Modeling Technique (OMT), which is for software development . However, these models and methodologies does not provide conceptual guidance for addressing specific challenges of AR interface design such as implicit style

Table 2 shows our view as to check whether the existing design models and methodologies meet the selective needs of AR interfaces. As the table indicates, none of the four design models and methodologies that we reviewed adequately meets the distinctive needs of the AR interfaces.

In spite the lack of design models and methodologies that comprehensively address the needs of AR interfaces, there is significant amount of useful work

Table 2. Characteristics of existing design models and methodologies

Design model and methodology	Support for distinctive characteristics of VR interfaces			
	Object graphics	Object dynamism	Communication patterns	Implicit interactions
Four level	No	No	Minimal	No
CLG	No	No	No	No
OAI	Yes	No	No	No
OO	No	Yes	Yes	No

in the AR field ,that can be used as building blocks in developing such models and methods. The scope and complexity of AR interfaces have been originated from ongoing efforts in the VR field to develop design frameworks , languages and other tools that can support the design of AR interfaces. These frameworks, languages and tools provides solution specific characteristics and requirements of VR interfaces such as the behavioral aspects, a particular issue within these domains or interaction aspects.

### Working Methodology to be Used

The following software with corresponding versions are used

- Unity 2017
- Vuforia 7
- Generate a license key for the application.
- Create a Database by adding the image target
- Download the specific dataset Download Dataset >> Unity editor''.
- Integrating with Unity.

### Challenges and Limitations

- Proper lightning condition are needed.
- The problem lies in other hardware as there are heaps of real time information which devices like mobile are not powerful enough to process.
- People doesn't rely on small screen devices for superimposing digital data like information on cell phones.
- There are some limitation of rendering digital data in some augment reality apps . some scaling issues are observed by changing the environment.
- Public unawareness.

### Expected Result

**Dynamic visiting card:-** For better visualisation of the person.

**3D object viewer:-** For scaling ,moving ,rotating the 3D object .

**Text recognition :-** Vuforia can detect words that belong to a pre-defined word list A list of 100,000 commonly used English words is what you get in SDK that you can use into your Text Recognition module. Custom word lists and filters can also be added to the block words from being recognized

**Image Classifier :-** multiclass object classification and recognition using smartphone and cloud computing (client server) technology.

## Conclusion

In this project we had proposed an app which will make visiting card more dynamic and interactive and ensures that viewer gets real time idea of the person . It gives user a new experience to immerse in augmented reality which is similar to reality

## Future Work

- Geo location and its live display of advanced maps. Augmented Reality providing suggestive and classified information to the end user (grocery stores, bars outlets ,restaurants, fashion).
- Industrial, military and medical applications A 3D analysis and view of the combat area in real time to tackle and deal with the problem
- In future we develop a hybrid software which will be accessible for all platform.

## Acknowledgments

We owe special thanks and gratitude debt to Professor

MUKESH KUMAR Department of Computer Science & Engineering,

IMS Engineering College, Ghaziabad for his regular guidance and support throughout the period of our work. His guidance have been in a very directive way .

## References

- *Imran I. Bohari Ashwini Singh. "image stitching". <https://www.overleaf.com/read/>*
- *Wikipedia contributors. "image stitching". [https://en.wikipedia.org/w/index.php?title=Image\\_stitching&oldid=719623757](https://en.wikipedia.org/w/index.php?title=Image_stitching&oldid=719623757), May 2016.*
- *Research on Image Registration Algorithm and Applications. Fudan University, Shanghai*
- *"[https://library.vuforia.com/articles/ Training/getting-started-with-vuforia-in-unity-2017-2-beta.html](https://library.vuforia.com/articles/Training/getting-started-with-vuforia-in-unity-2017-2-beta.html)*
- *<https://www.sitepoint.com/how-to-build-an-ar-android-app-with-vuforia-and-unity/>*
- *<https://blogs.unity3d.com/en/2018/01/15/vuforia-in-unity-build-cross-platform-ar-apps/>*

# Internet Protocol IPV5 Problems & Solutions Over IPv6

Mr. Sameer Asthana\*  
Khushboo\*\*

## ABSTRACT

The first widely used IP protocol was defined in RFC 791. RFC mean Request for Comments which describes overall behaviour, methods, and innovations that are applicable working with the Internet as well as Internet system. The IP address stands for "Internet Protocol". Where the "address" part of IP refers to a unique number that gets connected to all the online activity what you do. This is places on the Network Layer of the OSI reference Model. The network layer is responsible for the Hop-to-Hop transaction of packets and promised to do a reliable transaction of the packet to it's destination. There are number of phases of the IP address where some are successful and some of them are not in highlights. As IPv9 was also developed but as a April fool protocol by the IEFT again in 2004 china tried to develop it but it was not recommended as IP. Even though it was prediction that it would contain more additive functionality. And again IPv8&IPv12 launched but given no recommendation to it by IEFT.

### Introduction:

The Internet Protocol, is responsible for addressing, delivering and routing the packets that is sent by the sender over the network for its destination and also your online requests precisely. It includes an "electronic return address" to all your online requests and activity which is done for you. The address it uses is known as IP address to provide connection. Your computer is connected to the Internet. When you go online for sending or receiving email, to do shopping or chatting, your request has to be sent out to the desired destination anyhow, and the responses and information come back to you directly.

For this IP address plays a significant role. Generally IP emphasis on IPv4 & IPv6 but what about IPv1, IPv2, IPv3, IPv5? Previously the TCP was combined with the IP (TCP/IP). TCP was developed 3 times i.e. IPv1, IPv2, IPv3, known as TCP protocol. Again it was going to develop that was fourth time, that time Developers invent a new separate protocol IP that was IPV4. That's why IP start from IPv4. IPv4 contains 32 bit (4 octets) which feel not enough

address to provide a unique address in future. Because of this limitation developers thought to develop new IP standard, thought to develop IPv5. But developers declare it as a failed IP it's QOS (Quality of Service). And IPv5 was replaced by IPv6 with some new features.

### Internet Protocol (IP):

In 1970 the DOD (Department of Defence) developed TCP, to provide Transport Layer Function as well as Network Layer Function. But very soon it seems an inflexible solution and then it separates IP address with Network Layer.

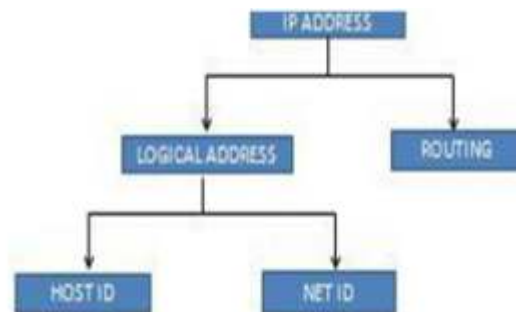


Figure 1

\*Asst.Prof. United College of Engineering & Research, Greater Noida., sam.asthana@gmail.com

\*\*M.Tech CS, United College of Engineering & Research, Greater Noida., Khushboo.alam05@gmail.com



**Logical Address:**

It provides unique address which includes Host Address and Network address.

**Routing:** Routers are devices which connect to two or more networks and they are combination. It provide best and short route to the packets. In RFC 760 the IP was originally defined, and has been revised number of times as well. IP Version 4

(IPv4 ) was the first version to experience, and is defined in RFC 791. IPv4 and IPV6 will be the focus of this guide.

**IPV4 (IP address version 4):**

The IPv4 protocol was invented in 1981, IPv4 includes a 32-bit address, And the number of possible IP addresses is  $2^{32}$  (4,294,967,296).

**IP Address Classes:**

- Class A:** (0-127) leading bit pattern is 0 (00000000)
- Class B:** (128–191) Leading bit pattern 10 (10000000)
- Class C:** (192 – 223) Leading bit pattern 110 (11000000)
- Class D:** (224–239) (Reserved for multicasting)
- Class E:** (240 – 255) (Reserved for experiments and research.  
32 bit

<b>Class A</b>	Net_ID	Host_ID	Host_ID	Host_ID
<b>Class B</b>	Net_ID	Net_ID	Host_ID	Host_ID
<b>Class C</b>	Net_ID	Net_ID	Net_ID	Host_ID
<b>Class D</b>	Net_ID	Net_ID	Net_ID	Net_ID

- Class A:** Num of Network installed is  $2^8$  Num of Host installed is  $2^{24}$
- Class B:** Num of Network installed is  $2^{16}$  Num of Host installed is  $2^{16}$
- Class C:** Num of Network installed is  $2^{24}$  Num of Host installed is  $2^8$
- Class D&E:** Num of Host installed is  $2^{32}$  (For experimental use)

**Subnet Mask :** Subnet mask identify which part of the address identifies network and which part of the address identify Host.

Subnet mask is a process to identify the Host number by dividing it into two parts.

Number of existing hosts on a subnet is  $2^{h-2}$ . Where h is the amount of bits used for the host segment of the address. The number of obtainable **subnets is  $2^n$** . Where n is the number of bits used for the network segment of the address.  
Such as 138.90.144.3 255.255.0.0  
This IP address has a subnet mask of 255.255.0.0.

**There are two rules of subnet mask:**

- 1) If a binary bit is set to 1 (or ON):Then corresponding bit in given address will identify the network address.
- 2) If a binary bit is set to a 0 (or off ) : Then corresponding bit in given address will identify the Host Address.

**Example IP Address:**

10111110.01110000.10101100.00110011  
Subnet Mask:  
11111111.11111111.00000000.00000000

The first 16 bits of the subnet mask will identify the network address because it set to 1. The last 16 bits of the subnet mask will identify the network address because it set to 0.

**Default Subnet Masks :**

<b>Class A</b>	<b>255.0.0.0</b>
<b>Class B</b>	<b>255.255.0.0</b>
<b>Class C</b>	<b>255.255.255.0</b>
<b>Class D &amp; E</b>	<b>255.255.255.255 (Multicasting)</b>

**Replacement of IPv4 to IPv6 Address:**

With the increasing number of new devices being connected to the Internet, That’s why it need to assigned new Address as well.

Due to limited Address space inventors find difficulty to allocate a unique address in future.

That’s why they decided to move on a new version of IP address i.e. **IPv6**.

**IPv6 (IP address version 6):**

The latest revision of the Internet Protocol (IP) is the **Internet Protocol version 6 (IPv6)**.

The communications protocol provides an identification and location system for computers over the networks and routes traffic across the Internet. IPv6 was invented by the **Internet Engineering Task Force (IETF)** to deal with problem of IPv4 address.

IPv6 is planned to replace IPv4, which still carries majority of Internet traffic over the network as of 2013.

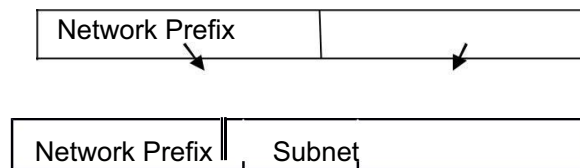
**Hardware Addressing:**

It identifies a host address within a local network. The 48 bit MAC addresses work as its hardware address and resides on Data Link Layer.

It uses 48 bit hexadecimal and 0-9, A-F combination of address of hardware. Such as 09:A2:43:BE:65:EF also can represent as 09A2:43BE:65EF

The First six hexadecimal of MAC address define the physical manufacturer and known as (OUI) i.e.

**ORGANIZATIONAL UNIQUE IDENTIFIER.**



Last six hexadecimal identifies the Host ID.

**ARP(Address Resolution Protocol):**

It is used to identify the logical address, through Physical address of the host over the network.

Every device over the Internet must have an IP address in order to communicate with other devices on the internet. IPv6 includes a 128-bit address (64 bit network address, 64 bit host address), allows 2<sup>128</sup>,

or around 3.4×10<sup>38</sup> addresses, or more than 7.9×10<sup>28</sup> times of IPv4.

IPv6 addresses represented as eight groups of four hexadecimal digits separated by colons or

terminators, for example 2011:0db5:86a3:0053:1012:8f3e:0650:7654.

**IPV6 Advantages:**

Very large Address Space (typically included host address of the end user prefixes.

It discovers its neighbour automatically in place of through ARP. (But still include the option for ARP and DHCP).

Hosts configured for autoconfig which is a default configuration on most of the platforms can automatically learn the prefix (es) and router(s) present on the segment, and can automatically address themselves as a appropriate using **EUI-64 addressing**.

Can have more clients on the subnet.

**IPV6 Limitations:**

It uses 128 bit address space that is some time unnecessary bit to transfer.

Ipv6 causes extra CPU cycles on a router/switch and has a higher bandwidth overhead.

Because IPv6 creates another stack of routing table during running than IPv4, that's why memory and CPU usage increased.

It required extra resources in any organization if an organization wants to use it.

It is not so much feasible with devices which use hardware build switching and routing quotes when dealing with IPv6.

**IPv6 is less efficient.**

IPv6 support from most vendors is immature at its best result.

Currently IPv6 exists quite a bit of overhead in network, and this is mainly because of from how IPv6 is encapsulated or tunnelled over IPv4.

File transfer performance in LAN between two server is almost same as IPv4.

Hard to remember the IP address.

undergone through experimental "Internet Stream Protocol" in 1979, which was never referred to as IPv5 and never introduced for public use.

#### **IPV5 Packets:**

IPv5 was designed to replace IPv4. IPv5 allows for a limited and flexible set of options that make the packet simple and very powerful. IPv5's main characters are that it can automatically sets up networks and new nodes in network and it allows almost 100% efficiency during periods of congestion. It contains 120 bit and assigned packets as 30 bit ,60 bit, 120 bit. The idea was that different-different computer can use small size of packet or according to its use. A packet header must be as small as possible, because that packet header will be sent with every packet on the internet, and every router has to support this packet header. IPv5 can support almost every addresses as IPv6 can, but IPv5 has ability to send 120-bit addresses in only 30 bits of data, other 90 bits are the same as its current network (and thus it passes through the target point i.e. destination and every router in the network).

IPv5 is designed in such a way that routers can and must set themselves in the network and configure a network automatically. This reduces workload of network administrators, and it allows networks to be easily created and to be maintained instead of any requirement of extensive set up, IPv5 allows routers which have few options that tells administrators how IPv4 installed machine cannot run IPv6.

Costly to replace IPv4 to IPv6.

Takes time to convert IPv4 to IPv6.

Turn off IPv4 before running IPv6.

#### **Difficult to memorize**

#### **Hard to place Prefixes on Topology Drawings:**

(Since it has long addressed so it become hard to overlay it onto the network topology. Even it much easy to do with IPV4. Even though we uses shorthand to remove the Zeros but still not usefull).

#### **IP Scanning is tough:**

As it has 264 possible host addresses per /64 prefix, it performs a ping scan to detect which devices is useless. Which increases security risk?

Why IPv4 to IPv6 Where is the IPV5 (IP address Version5)?

Most widely uses IP address is IPv4 and IPv6, but questions arises where is IPv5. The number 5 was

to control the network through router. This system permits faultless mixing of networks by using other protocols inside an IPv5 network. Another network will look like one single system by using many IP addresses like any other network. This also can allow some very special applications like ad-hoc wireless networks to use different network protocol that optimizes for power rather than speed.

The protocol is built in such a way so that very few bits are required for the header. This can be significant in wireless protocols. Because in wireless network every bit requires not only bandwidth but also power, which is a limited resource in wireless applications. The protocol is also built so that routers need to accomplish its task or operations very fast. It was designed to be smooth and powerful with manageable array of options that allows end systems to modify what services they need, and what services they don't need. Every packet are much capable to select a priority, and make a real-time exchanges smoother. Such as ATM". It allows less priority packets to give up to much urgent packets i.e. making real - time exchanges smoother.. Every packet can contain the IP compatible version field by itself.

**Design of IPv5 :** While designing IPv5 it consider some points i.e. Minimum delay, maximum throughput, packet ordering, congestion control, corruption, correction, security, least-cost routing, routing information propagation, automatic configuration and administrative autonomy. It also consider needs like that of real-time applications and broadcast media.

**Type 0 packet: bare-connection**

This type of packet bare minimum for the packet they are sleek and can be processed as fastest.

The router table are fix and not able to make any change in packet priority .Packet priority cannot be change in IPv5 Also the TTL (Time to live ) is dictated by the priority and have no right to make any change.

The bare type packet are smooth and processed fastest but they can dropped if any oversized , or congestion in the network, thus it increases a loss rate and error rate.

the congestion level which is experienced by the other router in the network.

**Type-3:** In type-3 routers require reliable loss rate which can be used to verify that whether the packet was received, or not or to resend if necessary. It allows routers to set packets to a type 3 reliable loss rate, and by this network can approach 100% theoretical efficiency in times of congestion and does not effect the efficiency. As it does in IPv4. IPv5 wants the packet header minus the lower 3 bits of the Time-to-live, and minus payload ,to have an error rate of 0.01% despite of the reliable error rate explained in the packet. Packet forwarding can agree to end systems to identify part of the route to the destination, which could be helpful for some security reasons. Packets can tell a receiver that there is some loss of data than it was originally sent .so it could be allowed as a packet to be truncated (rather than fragmented) if it become too large for a particular router otherwise Multicasting is a great option to send same data to 256 different recipients . Since a multicast packet is split onto routers based where the addresses depart, this option has the capability to reduce the total load on both the network as well as the sending system.

**It contain priority of 2 bit of three types priority:**

1. It could be sent out at most 1/4 times as frequent as priority 1 packets. Which is Even TTL are not able to identify that how long a packet will linger on the network.

IPV5 header does not contain any segment f data, that's why it become difficult to recognize the header information. It contains much smaller overhead bits than an IPv4 packet. 66 bits are minimum overhead if using the DTP header.

**Type-1:** This type of packets are extremely flexible even though it has a few points of variability. The TTL (Time to Live) field goes back to check time to kill a packet, which provide to the user the ability to know the estimated time to wait that how much time the packet will be in network. As IPv5 packet's header doesn't include any error checking information or sequence numbers information. So that different link-layers may have different essential error rates and packet-loss rates. IPv5 only specifies the required facilities and permits the link-layer to provide these services with several extra functionality which is necessary.

**Type-2:** It is Link- to-link routing information ,which contain 4 bit as one i.e. (0101). The packet type is of 2 bit and can include bare-connection minimum end-to-end header size , normal end-to-end packet, routing information. In type-2 the routing may contain the overall information about proposed for background traffic and control - e.g. news, E-mail, ICMP, SNMP, Routing Information and so on.

- 2. It is sent out between 2 and 4 times as repeatedly or frequently as priority 0 packets. Which is intended to unattended data transfer as well as small real-time traffic for e.g. Torrents.
- 3. It is sent out between 4 and 8 times as commonly as priority 0 packets. Which is intended for attended mass data transfer and medium-sized real-time traffic for e.g. Streaming, html.
- 4. It is sent out between 32 and 64 times as frequently as priority 0 packets which is required for bulk real-time traffic -For e.g. video/audio chats, online games, remote Login and so on.

It contain 16 bit of packet length and to get the full packet-size from this , add 105 bits (which is the size of the minimal packet).

The Time-to-Live includes 6-bit and each "tic" of the TTL field represents an amount of time related to the priority. If it is less than the time i.e. given by the TTL size, then the TTL field is probably reduced by one, by this it is decremented by one fourth of the routers through which it travels.

It can send Jumbo packet of 1 bit of two types priority i.e. 0 & 1.

0: Packet length includes a normal number of bytes

whereas priority of

1. Packet length includes the value of the field times  $2^{15}$

**The Reliable Loss Rate is of 2-bit:**

The most importantly the link-level service demanded for packet arrival to the destination.

0: It guaranteed maximum avg link-to-link loss rate i.e. of 40%

1. It guaranteed maximum avg link-to-link loss rate i.e. of 10%
2. It guaranteed maximum avg link-to-link loss rate i.e. of 1%
3. It guaranteed maximum avg link-to-link loss rate i.e. of 0%. It is required to use sequence numbers and sequence checking, AWKs/NAKs (Acknowledgment / Negative Acknowledgment), packet buffering and packet retransmission on NAK. Through this packets are automatically switched to this level of reliability by routers during network congestion.

forward IP 2 and what is now in forward IP 1, etc. for the security reason.

Truncated : If the packet is too big for a router in the path.

Multicast Type:

- On 0: It does normal multicast
- On 1: It does group-cast

IPv5 Protocols: IPv5 authorized routers that must support some different services which includes address allocation, packet forwarding, multicasting, and of course routing. A router must support at least 10,000 bytes of a packet length. This provision ensures that the maximum packet length of  $25^3 \cdot 323$  bits is supported, along with a margin for a fair sized of minimum payload. A 72 bits addresses are long and are written similarly to IPv6 addresses, but uses five sets of three digits in base-

32 lower-case which is known as tria-contakai-decimal notation separated by dots. Which is very similar to the double dot like double colon in IPv6 that can be used to specify a string of zeros.

Network structure and Address allocation:

The addresses are allocated by themselves in a hierarchical fashion and networks can be set up

A type 3 reliable loss rate is a loss rate of 0% otherwise the router search that it has no longer a link to the destination, in this case the packet must be drop out.

The reliable Error Rate of 2-bit:

The most importantly link-level service demanded for the rate of bit errors in the packet's payload is the reliable error rate.

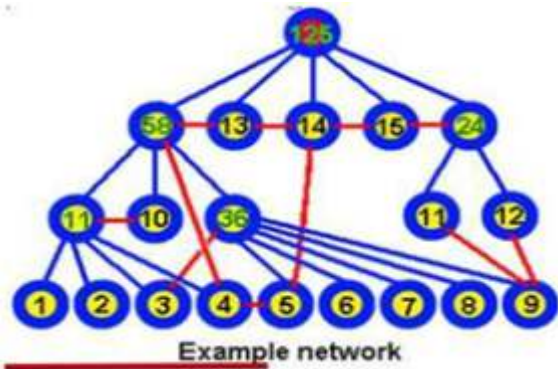
- 0: It guaranteed maximum average link-to-link error rate i.e. of 20%
- 1: It guaranteed maximum average link-to-link error rate i.e. of 5%
- 2: It guaranteed maximum average link-to-link error rate i.e. of 1%
- 3: It guaranteed maximum average link-to-link error rate i.e. of 0.01%

Packet Forwarding: when a packet reaches to its destination, that time its destination is changed to forward IP 1 and sent back out but before the packet has to sent back out, the Forwards field is decremented, and the router become certain not to send forward IP 1 out.

It does to be used for security purposes essentially removing packet from the queue that what was in



independently, and then linked together later without any pre-configuration.



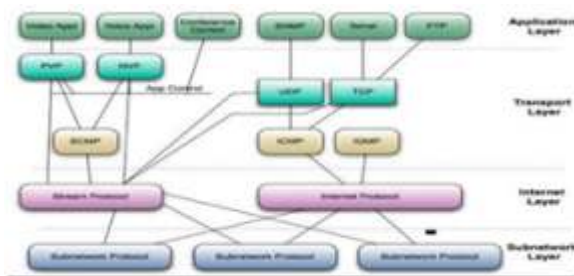
**Parent node connection**

The parent node is responsible for allocating and changing the IP address of its children node and verifying the routing information like what the identity of routers connected to the child through shortcuts or parent node connections.

The router that can support more addresses will become the parent of the other router where a parent-less router connects to another parent-less router, but the network of the router if contain the

least allocated addresses that will be the one to change IP addresses.

**IPV5 Architecture:**



**Application Layer:** This is the seventh layer of the OSI\_Model on which a user work.

**Internet Layer:** It is the third layer of the OSI\_Model and responsible to allocate the IP address to the packets. It includes Stream Protocol as well as Internet Protocol.

**Sub network Layer:** It is the Layer between Internet layer and Data Link layer.

**Limitations:**

In IPv5 Sometimes efficient operations which require that the lines exist between internet abstraction layers are blurred, and such blurring causes the sake of efficiency. Reducing efficiency on a universal scale for conceptual ease and interoperability between layers. Which is neither

appropriate nor realistic. If a private user sending a top priority non real time packets, or any real time top priority packets, so the user will experience extra delay. It provides fast and almost congestion free data delivery but due to network interface blurring that affects the efficiency. There is a drawback for the private user who requires upper-priority packets. If a user wants to transfer non-real-time packets which are labelled as top priority, then any real-time top-priority packets user will experience extra delay. This could be resolved or avoided if it had the non-real-time packets which are labelled with the correct priority. Routers are neither free to modify the packet priorities nor free to make any complaint to IPv5.

**Advantages:**

**Over IPv6:** Uses 120 bit where all bits are useful but in IPv6 sometime 128 bit become extra bit address.

**Over IPv4:** IPv4 degrades the efficiency if reliable loss rate is not accurately used by router.

**Solutions:**

The Network Technology can use to remove Blur to the network interfaces. ongestion control, Error controls (CRC) but provide alternative by using ATM. Dynamic router should use to give priority to the packets or again giving another flexibility to overcome this. As IPv5 is put in the Internet streaming protocol and it has not given the name of "Internet Protocol-version 5" by IEFT.

**Internet Stream Protocol:**

The Internet Stream Protocol (ST and later ST-II) is known as family of experimental protocols which is defined in "Internet Experiment Note" IEN-119 (1979), and revise in RFC 1190 (Request for comment).

The Internet Stream Protocol family never announced publicly for the public use, but number of concepts are available and even using in ST that are very similar to later technologies. Such as "Asynchronous Transfer Mode" protocols. It can be found in "Multiprotocol Label Switching (MPLS)". ST arose as the transport protocol of the Network Voice Protocol

**Conclusion:**

IPv5 was not successful and the IEFT does not given name to it as IPv5 because they think that it does not provide the quality of services. Even the organization like IBM and all which are using this they have already given it the name of "Magical 5". So it is not the failed experimental protocol i.e. IPv5.

Even it need some focus on its working and some revised work.

Due to some limitation it was not so popular. But yes no doubt it is a "Magical -5".

**References:**

- [1] Henry Chukwuemeka Paul, Kinn Abass Bakon, "THE IMPORTANCE OF THEIR CO-EXISTENCE A STUDY ON IPv4 and IPv6: THE IMPORTANCE OF THEIR CO-EXISTENCE A STUDY ON IPv4 and IPv6: THE IMPORTANCE OF THEIR CO-EXISTENCE" *International Journal of Information System and Engineering*, Vol. 4 (No.2), November, 2016.
- [2] Aluko T.S, Olusanya O.J, Oloyede O.E, Ebisin A.F "Comparative Analysis between Internet Protocol Version 4 & 6 (IPv4 and IPv6) " *International Journal of Scientific & Engineering Research*, Volume 5, Issue 8, August-2014.
- [3] Arathi Ramani, Salim Vhora, S. Sanyal, " The Next Generation Internet Protocol".
- [4] <http://playground.sun.com/pub/ipng/html>
- [5] NGtrans: <http://www.6bone.net/ngtrans>
- [6] IPv6 users site: <http://www.ipv6.org>



# FEEDBACK FORM

Your valuable comments will help us to shape the future issues better

	Highly Appreciable	Somewhat Appreciable	Not Appreciable	Did Not Read
Energy Efficiency Of Cloud Computing: A Review of The Recent Advances	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
A Review on Identification of Fake Profiles in Online Social Network	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Title: A Step Towards Protecting Innocent Civilian of Internet Terrene	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Security For Sap Erp System - A Case Study	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Movie Recommender System By Genetic Algorithm	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Image Classification Through Convolutional Neural Network	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Big Data & Hadoop: Characteristics And Importance	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Information Security & Cyber Crime: An Inside Story of threats, lacking & solution	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Use Of Augmented Reality To Build Various Application Making World More Interactive and Informative	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Internet Protocol Ipv5 Problems & Solutions Over Ipv6	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**Comments/Suggests(if any):** \_\_\_\_\_

**Name :Mr./Ms.** \_\_\_\_\_

**Designation:** \_\_\_\_\_ **Organization/Institution:** \_\_\_\_\_

**Address:** \_\_\_\_\_

**Phone:** \_\_\_\_\_ **E-mail:** \_\_\_\_\_

## SUBSCRIPTION FORM

I wish to subscribe to / renew my subscription to " IPEM Journal of Computer Application & Research" for 1 / 2 / 3 years(s).  
 A bank draft /cheque\* bearing no.....dated.....drawn in favour of INSTITUTE OF PROFESSIONAL EXCELLENCE & MANAGEMENT, payable at GHAZIABAD / DELHI towards subscription for ..... years, is enclosed.

**Name:** .....

**Org. / Ins.:** .....

**Address:** .....

**City:** ..... **Pin**.....

**Phone:** .....

**Fax:** .....

**Mobile:** .....

**E-mail:** .....

**Category:** .....

**Year:** .....

Subscription Rates			
Category	1yr.	2yr.	3yr.
Indian (in Rs.)			
1. Institutions	300	550	800
2. Individual	200	350	550
3. IPEM Student/ Alumni	150	250	350

\_\_\_\_\_  
Signature with date

---

**The Editor**

**The IPEM Journal of Computer Application & Research  
Institute of Professional Excellence & Management  
A-13/1, S.S. G.T Road,  
Industrial Area, NH-24 By Pass  
Ghaziabad-201010  
Tel.: 0120-4174500**

Affix  
Postal  
Stamp

---

**The Editor**

**The IPEM Journal of Computer Application & Research  
Institute of Professional Excellence & Management  
A-13/1, S.S. G.T Road,  
Industrial Area, NH-24 By Pass  
Ghaziabad-201010  
Tel.: 0120-4174500**

Affix  
Postal  
Stamp



---

## GUIDELINES FOR AUTHORS

All papers are subjected to a blind peer review process. Manuscripts are invited from Academicians, Scientists and Research Scholars for publication consideration. Papers are accepted for editorial consideration through email [journalit@ipemgzb.ac.in](mailto:journalit@ipemgzb.ac.in) with the understanding that they have not been published, submitted or accepted for publication elsewhere in the same form, either in the language of the paper or any other language without the consent of the Editorial Board.

**Title :** The title of the paper should be concise and definitive.

**Length :** Contributions should not exceed 5000 words including Charts, Tables and other Graphics. The order of the paper should be preceded by an Abstract (200-300word) ,Introduction, main text, tables and charts in black & white only(with caption) and figure captions, list of symbols and abbreviations (if necessary), conclusion and list of references.

**Abstract :** A short abstract (200-300) words should give a clear indication of the objective, scope, and results of the paper. Some Key words must be provided.

**Authors Names and Affiliation:** It is a journal policy that all those who have participated significantly in the technical aspects of a paper be recognized as co-authors or cited in the acknowledgements. A cover page containing the title, author's name & affiliation, mailing address including city, state, and zip code, mobile number, fax number and e-mail address.

**Copyright:** Each manuscript must be accompanied by a statement that it has been neither published nor submitted elsewhere for publication, in whole or in part.

**Paper Acceptance:** The final decision on publication is made by the Editor-in-Chief upon recommendation of Editorial Board members. Acceptance of papers for publication shall be informed through e-mail or through normal mail. Manuscripts that fail to conform to the guidelines will not be considered for publication. The author whose paper is published will receive one free copy of journal that carries the paper.

**References:** The references should be brought at the end of the article, and numbered in the order of their appearance in the paper. References should include full details of the name(s) of the author(s), title of the article or book, name of the journal, details of the publishers, year & month of publication including page numbers, as appropriate. References should be cited in accordance with the following example:

- 1 Mishra, K.M. (2002) Knowledge Management, New Delhi: Pearson Education MA: Allyn & Bacon.
- 2 Rowling, J.K. (2001) Harry Potter and the Socerer's Stone. London: Bloomsburg Children's.
- 3 Tyagi, R.M, and Malik, S.P. (2007) Job Satisfaction Working Paper No 46, Indian Institute of Travel Management, Gwalior.
- 5 Jacoby, W. G. (1994). Public attitudes toward government spending. American Journal of Political Science, 38(2), 336-361. Retrieved from <http://www.jstor.org>.

**Publication Cost:** Free Publications \ No Page Charge.

**Frequency of Publication:** Annual

---

## About the Institute

IPEM made a modest beginning in the year 1996, with few Management and Computer Application Programmes. Today the IPEM Group of Institutions are in the forefront of imparting knowledge in the field of Education, Law, Management and Information Technology. The Computer & IT department was started in 1996 with Bachelor of Computer (BCA), affiliated to the Chaudhary Charan Singh University, Meerut with 120 seats. This journal of Computer & IT Department students are exposed to emerging trends in the areas of information Technology by value additions through workshops, Live Project and a regular interaction with Experts from Industry. This is reflected in the performance of the students as we have 100% result with maximum 1st division. We provide best placement to the students.

The Computer & IT Department is running two courses successfully: Master of Computer Application (MCA) is approved by All India Council for Technical Education (AICTE) and affiliated to Dr. A.P.J. Abdul Kalam Technical University (APJAKTU) Lucknow and Bachelor of Computer Application (BCA) is affiliated to the Chaudhary Charan Singh University, Meerut. The other courses are running under IPEM group of Institution are Master of Business Administration (MBA) and Master of Applied Management (MAM) approved by all India Council for Technical Education (AICTE) and affiliated to DR. A.P.J. Abdul Kalam Technical University (APJAKTU) Lucknow. The Post Graduate Diploma in Management (PGDM) is approved by All India Council for Technical Education (AICTE) Govt. of India, Ministry of HRD. The Bachelor of Business Administration (BBA), Bachelor of Law (LLB) (3 years) BALLB (5 Years) approved by Bar Council of India and affiliated to the Chaudhary Charan Singh University, Meerut, Bachelor of Education (B.Ed.) and Basic Teacher Certificate (BTC) approved by National Council for Teacher Education (NCTE). Bachelor of Education (B.Ed.) is affiliated to the Chaudhary Charan Singh University, Meerut and Basic Teacher Certificate (BTC) is affiliated to the State Council of Educational Research and Training (SCERT) Lucknow.

The focus of IPEM has always been to be at the forefront of optimum utilization of IT resources and leverage the power of IT in making the learning process, informative and engaging. The students are provided with hands on experience and learning process, informative and engaging. The students are provided with hands on experience and learning with the state-of-the-art technology.

The computer & IT Department has enriched with well equipped labs in Aryabhata Block i.e. Programming Lab for the specialization in Database, Java, .Net etc (Aryabhata Lab-1), Internet Lab (Aryabhata Lab-2) and for UNIX, LINUX, Android etc (Aryabhata Lab-3). The Computer & IT Department of IPEM group of Institutions prepare the students who would be able to lead the future Industry and chase the world-wide mega trends. The Department has shined covered out for itself's commanding position with best results and placement.

IT Department of IPEM Group of Institutions organizes various workshop and seminar on latest IT trends every year. Seminars often feature several speakers, each one providing information from a different angle or perspective. People who attend seminars learn new ideas and skills to help them improve their production, while those who present at seminars gain exposure for their products or services. Presenting at an academic seminar is an important part of a researcher's/Scholars life, and is an opportunity that most young researchers look forward to. A good mix of paper presentations and journal publications is important when looking to move up the academic career ladder as well.

Spacious Lecture Theaters are thoughtfully designed to induce high quality learning and are equipped with high and teaching aids such LCD and OHP projectors. Priority is attached to achieve optimal convergence of stimulating pedagogy & enabling environment. The latest audio-visual aids and multimedia technology enables the Faculty members to have interactive sessions. Classroom learning is meant primarily for theoretical and conceptual input & consolidated by combining lectures with Case methods and Group Discussion for group learning. Extensive use of laptops is made by students in the well networked class rooms.

