BIG DATA

Abstract

Big Data is a latest term introduced to define large and complex datasets. Due to their size and complexity, it is not possible to manage them with our conventional techniques or data mining tools. Using Big Data mining organizations can extract useful information from these large pool of data or streams of data. By analysis of this datasets useful statistics can be extracted. In spite of the usability of Big Data, there are several challenges related to it. This challenge is becoming most evolutionary area of research for the coming years. The paper presents an overview of the topic, methodologies and forecast to the future. Big Data concern large-volume, complex, growing data sets with multiple, autonomous sources. With the fast development of networking, data storage, and the data collection capacity, Big Data are now rapidly expanding in all science and engineering domains, including physical, biological and biomedical sciences. This paper presents a HACE theorem that characterizes the features of the Big Data revolution, and proposes a Big Data processing model, from the data mining perspective. This data-driven model involves demand-driven aggregation of information sources, mining and analysis, user interest modeling, and security and privacy considerations. We analyze the challenging issues in the data-driven model and also in the Big Data revolution.

1. Introduction

Our capability for data generation has never been so powerful and enormous ever since the invention of the Information Technology in the early 19th century. As another example, on October 4, 2012, the first presidential debate between President Barack Obama and Governor Mitt Romney triggered more than 10 million tweets within two hours (Twitter Blog 2012). Among all these tweets, the specific moments that generated the most discussions actually revealed the public interests, such as the discussions about Medicare and vouchers. Such online discussions provide a new means to sense the public interests and generate feedback in real-time, and are mostly appealing compared to generic media, such as radio or TV broadcasting. Another example is Flickr, a public picture sharing site, which received 1.8 million photos per day, on average, from February to March 2012 (Michel F. 2012). Assuming the size of each photo is 2 megabytes (MB), this resulted in 3.6 terabytes (TB) storage every single day. As “a picture is worth a thousand words”, the billions of pictures on Flicker are a treasure tank for us to explore the human society, social events, public affairs, disasters etc., only if we have the power to harness the enormous amount of data. The above examples demonstrate the rise of Big Data applications where data collection has grown tremendously and is beyond the ability of commonly used software tools to capture, manage, and process within a “tolerable elapsed time”. The most fundamental challenge for the Big Data applications is to explore the large volumes of data and extract useful information or knowledge for future actions. In many situations, the knowledge extraction process has to be very efficient and close to real-time because storing all observed data is nearly infeasible. the data generated from the SKA is exceptionally large. Although researchers have confirmed that interesting patterns, such as transient radio anomalies can be discovered from the SKA data, existing methods are incapable of handling this Big Data. As a result, the unprecedented data volumes require an effective data analysis and prediction platform to achieve fast-response and real-time classification for such Big Data.

2. Big Data Characteristics: HACE Theorem HACE Theorem: Big Data starts with large-volume, heterogeneous, autonomous sources with distributed and decentralized control, and seeks to explore complex and evolving relationships among data. These characteristics make it an extreme challenge for discovering useful knowledge from the Big Data. In a naïve sense, we can imagine that a number of blind men are trying to size up a giant elephant which will be the Big Data in this context. The goal of each blind man is to draw a picture (or conclusion) of the elephant according to the part of information he collected during the process. Because each person's view is limited to his local region, it is not surprising that the blind men will each conclude independently that the elephant "feels" like a rope, a hose, or a wall, depending on the region each of them is limited to. To make the problem even more complicated, let's assume that (a) the elephant is growing rapidly and its pose also changes constantly, and (b) the blind men also learn from each other while exchanging information on their respective feelings on the elephant. Exploring the Big Data in this 4 scenario is equivalent to aggregating heterogeneous information from different sources (blind men) to help draw a best possible picture to reveal the genuine gesture of the elephant in a real-time fashion. Indeed, this task is not as simple as asking each blind man to describe his feelings about the elephant and then getting an expert to draw one single picture with a combined view, concerning that each individual may speak a different language (heterogeneous and diverse information sources) and they may even have privacy concerns about the messages they deliberate in the information exchange process.

2.2 Autonomous Sources with Distributed and Decentralized Control Autonomous data sources with distributed and decentralized controls are a main characteristic of Big Data applications. Being autonomous, each data sources is able to generate and collect information without involving (or relying on) any centralized control. This is similar to the World Wide Web (WWW) setting where each web server provides a certain amount of information and each server is able to fully function 5 without necessarily relying on other servers. On the other hand, the enormous volumes of the data also make an application vulnerable to attacks or malfunctions, if the whole system has to rely on any centralized control unit. For major Big Data related applications, such as Google, Flicker, Facebook, a large number of server farms are deployed all over the world to ensure nonstop services and quick responses for local markets. Such autonomous sources are not only the solutions of the technical designs, but also the results of the legislation and the regulation rules in different countries/regions. More specifically, the local government regulations also impact on the wholesale management process and eventually result in data representations and data warehouses for local markets.

2.3 Complex and Evolving Relationships While the volume of the Big Data increases, so do the complexity and the relationships underneath the data. In an early stage of data centralized information systems, the focus is on finding best feature values to represent each observation. This is similar to using a number of data fields, such as age, gender, income, education background etc., to characterize each individual. This type of sample-feature representation inherently treats each individual as an independent entity without considering their social connections which is one of the most important factors of the human society. People form friend circles based on their common hobbies or connections by biological relationships. Such social

connections commonly exist in not only our daily activities, but also are very popular in virtual worlds. For example, major social network sites, such as Facebook .The correlations between individuals inherently complicate the whole data representation and any reasoning process. In the sample-feature representation, individuals are regarded similar if they share similar feature values, whereas in the sample-feature-relationship representation, two individuals can be linked together (through their social connections) even though they might share nothing in common in the feature domains at all. In a dynamic world, the features used to represent the individuals and the social ties used to represent our connections may also evolve with respect to temporal, spatial, and other factors. Challenges on information sharing and privacy, and Big Data application domains and knowledge form Tier II, which concentrates on high level semantics, application domain knowledge, and user privacy issues. The outmost circle shows Tier III challenges on actual mining algorithms.

3. Data Mining Challenges with Big Data

For an intelligent learning database system (Wu 2000) to handle Big Data, the essential key is to scale up to the exceptionally large volume of data and provide treatments for the characteristics featured by the aforementioned HACE theorem. The challenges at Tier I focus on data accessing and actual computing procedures. Because Big Data are often stored at different locations and data volumes may continuously grow, an effective computing platform will have to take distributed large-scale data storage into consideration for computing. example, while typical data mining algorithms require all data to be loaded into the main memory, this is becoming a clear technical barrier for Big Data because moving data across different locations is expensive (e.g., subject to intensive network communication and other IO costs), even if we do have a super large main memory to hold all data for computing. The challenges at Tier II center around semantics and domain knowledge for different Big Data applications. Such information can provide additional benefits to the mining process, as well as add technical barriers to the Big Data access (Tier I) and mining algorithms (Tier III). For example, depending on different domain applications, the data privacy and information sharing mechanisms between data producers and data consumers can be significantly different. Sharing sensor network data for applications like water quality monitoring may not be discouraged, whereas releasing and sharing mobile users' location information is clearly not acceptable for majority, if not all, applications. In addition to the above privacy issues, the application domains can also provide additional information to benefit or guide Big Data mining algorithm designs. For example, in market basket transactions data, each transaction is considered independent and the discovered knowledge is typically represented by finding highly correlated items, possibly with respect to different temporal and/or spatial restrictions. In a social network, on the other hand, users are linked and share dependency structures. The knowledge is then represented by user communities, leaders in each group, and social influence modeling etc. Therefore, understanding semantics and application knowledge is important for both low-level data access and for high level mining algorithm designs. At Tier III, the data mining challenges concentrate on algorithm designs in tackling the difficulties raised by the Big Data volumes, distributed data distributions, and by complex and dynamic data characteristics. The circle at Tier III contains three stages. Firstly, sparse, heterogeneous,

uncertain, incomplete, and multi-source data are preprocessed by data fusion techniques. Secondly, complex and dynamic data are mined after pre-processing. Thirdly, the global knowledge that is obtained by local learning and model fusion is tested and relevant information is fed back to the pre-processing stage. Then the model and parameters are adjusted according to the feedback. In the whole process, information sharing is not only a promise of smooth development of each stage, but also a purpose of Big Data processing. 8 In the following, we elaborate challenges with respect the three tier framework .Big Data Mining Platform In typical data mining systems, the mining procedures require computational intensive computing units for data analysis and comparisons. A computing platform is therefore needed to have efficient access to, at least, two types of resources: data and computing processors. For small scale data mining tasks, a single desktop computer, which contains hard disk and CPU processors, is sufficient to fulfill the data mining goals. Indeed, many data mining algorithm are designed to handle this type of problem settings. The role of the software component is to make sure that a single data mining task, such as finding the best match of a query from a database with billions of samples, is split into many small tasks each of which is running on one or multiple computing nodes. For example, as of this writing, the world most powerful super computer Titan, which is deployed at Oak Ridge National Laboratory in Tennessee, USA, contains 18,688 nodes each with a 16-core CPU. Such a Big Data system, which blends both hardware and software components, is hardly available without key industrial stockholders' support. In fact, for decades, companies have been making business decisions based on transactional data stored in relational databases. Big Data mining offers opportunities to go beyond their relational databases to rely on less structured data: weblogs, social media, email, sensors, and photographs that can be mined for useful information. Major business intelligence companies, such IBM, Oracle, Teradata etc., have all featured their own products to help customers acquire and 9 organize these diverse data sources and coordinate with customers' existing data to find new insights and capitalize on hidden relationships.

3.1 Big Data Semantics and Application Knowledge Semantics and application knowledge in Big Data refer to numerous aspects related to the regulations, policies, user knowledge, and domain information. The two most important issues at this tier include (1) data sharing and privacy; and (2) domain and application knowledge. The former provides answers to resolve concerns on how data are maintained, accessed, and shared; whereas the latter focuses on answering questions like "what are the underlying applications ?" and "what are the knowledge or patterns users intend to discover from the data ?".

3.1.1 Information Sharing and Data Privacy Information sharing is an ultimate goal for all systems involving multiple parties (Howe et al. 2008). While the motivation for sharing is clear, a real-world concern is that Big Data applications are related to sensitive information, such as banking transactions and medical records, and so simple data exchanges or transmissions do not resolve privacy concerns For the first approach, common challenges are to design secured certification or access control mechanisms, such that no sensitive information can be misconducted by unauthorized individuals. For data anonymization, the main objective is to inject randomness into the data to ensure a number of privacy goals. blood glucose level is clearly a better feature than body mass in diagnosing Type II diabetes). The domain and application knowledge can

also help design achievable business objectives by using Big Data analytical techniques

3.2 Mining Complex and Dynamic Data The rise of Big Data is driven by the rapid increasing of complex data and their changes in volumes and in nature (Birney 2012). Documents posted on WWW servers, Internet backbones, social networks, communication networks, and transportation networks etc. are all featured with complex data. While 13 complex dependency structures underneath the data raise the difficulty for our learning systems, they also offer exciting opportunities that simple data representations are incapable of achieving. For example, researchers have successfully used Twitter, a well-known social networking facility, to detect events such as earthquakes and major social activities, with nearly online speed and very high accuracy. In addition, the knowledge of people's queries to search engines also enables a new early warning system for detecting fast spreading flu outbreaks Making use of complex data is a major challenge for Big Data applications, because any two parties in a complex network are potentially interested to each other with a social connection. Such a connection is quadratic with respect to the number of nodes in the network, so a million node network may be subject to one trillion connections. For a large social network site, like Facebook, the number of active users has already reached 1 billion, and analyzing such an enormous network is a big challenge for Big Data mining. If we take daily user actions/interactions into consideration, the scale of difficulty will be even more astonishing. Inspired by the above challenges, many data mining methods have been developed to find interesting knowledge from Big Data with complex relationships and dynamically changing volumes Complex intrinsic semantic associations in data: news on the Web, comments on Twitter, pictures on Flicker, and clips of video on YouTube may discuss about an academic award-winning event at the same time. There is no doubt that there are strong semantic associations in these data. Mining complex semantic associations from "text-image-video" data will significantly help improve application system performance such as search engines or recommendation systems. However, in the context of Big Data, it is a great challenge to efficiently describe semantic features and to build semantic association models to bridge the semantic gap of various heterogeneous data sources. Complex relationship networks in data: In the context of Big Data, there exist relationships between individuals.. The size or complexity of the Big Data, including transaction and interaction data sets, exceeds a regular technical capability in capturing, managing, and processing these data within reasonable cost and time limits. In the context of Big Data, real-time processing for complex data is a very challenging task.

Conclusions

Driven by real-world applications and key industrial stakeholders and initialized by national funding agencies, managing and mining Big Data have shown to be a challenging yet very compelling task. While the term Big Data literally concerns about data volumes, our HACE theorem suggests that the key characteristics of the Big Data are (1) huge with heterogeneous and diverse data sources, (2) autonomous with distributed and decentralized control, and (3) complex and evolving in data and knowledge associations. Such combined characteristics suggest that Big Data requires a "big mind" to consolidate data for maximum values (Jacobs 2009). In order to explore Big Data, we have analyzed several challenges at the data, model, and system levels. To support Big Data mining, high

performance computing platforms are required which impose systematic designs to unleash the full power of the Big Data. At the data level, the autonomous information sources and the variety of the data collection environments, often result in data with complicated conditions, such as missing/uncertain values. In other situations, privacy concerns, noise and errors can be introduced into the data, to produce altered data copies. Developing a safe and sound information sharing protocol is a major challenge. At the model level, the key challenge is to generate global models by combining locally discovered patterns to form a unifying view. This requires carefully designed algorithms to analyze model correlations between distributed sites, and fuse decisions from multiple sources to gain a best model out of the Big Data. At the system level, the essential challenge is that a Big Data mining framework needs to consider complex relationships between samples, models, and 22 data sources, along with their evolving changes with time and other possible factors. A system needs to be carefully designed so that unstructured data can be linked through their complex relationships to form useful patterns, and the growth of data volumes and item relationships should help form legitimate patterns to predict the trend and future. We regard Big Data as an emerging trend and the need for Big Data mining is arising in all science and engineering domains. With Big Data technologies, we will hopefully be able to provide most relevant and most accurate social sensing feedback to better understand our society at real-time. We can further stimulate the participation of the public audiences in the data production circle for societal and economical events. The era of Big Data has arrived. Acknowledgements This work is supported by the National 863 Program of China (2012AA011005), the National 973 Program of China (2013CB329604), the National Natural Science Foundation of China (NSFC 61229301, 61273297, and 61273292), the US National Science Foundation (NSF CCF-0905337), and the Australian Research Council (ARC) Future Fellowship (FT100100971). The authors would like to thank the anonymous reviewers for their valuable and constructive comments on improving the paper.

References

[1] Algorithms for mining the evolution of conserved relational states in dynamic networks

[2] Knowledge and Information Systems, Jerome P. Reiter: Big privacy: protecting confidentiality in big data.

[3] Analyzing collective behavior from blogs using swarm intelligence, Knowledge and Information Systems,.

Cloud Computing – A New Perspective

Analyzing Cloud Computing on key Parameters

Amit Kumar
AWS Certified
IINTM
New Delhi, India
amkr95@gmail.com

Tomesh
IINTM
New Delhi, India
Tomeshkumar2017890
@gmail.com

Achint Sharma
IINTM
New Delhi, India

***Abstract*— In the recent time, cloud computing was developed and soon it became a very common and vital term of in IT industry, and we find everyone quite very open ear and also are talking about the same.**

But with everything some questions arise; now the question is; is this that good and beneficial to talk about or use? As people are talking about it, so there comes the difference and the inclination towards talking is much, and the users are less. It is kind of people are very reluctant to use & know. The cloud computing is classified or categorized further into 3 categories namely – Public Cloud, Private Cloud & Hybrid Cloud, the basis of classification was the reach (accessibility); of how soon or fast it can be reached, and resource location. We will now more on towards a more detailed description on cloud computing in terms of adaption feasibility, economic suitability, usage degree (easy, medium or hard), security and its advantages and disadvantages.

***Keywords*— Cloud Computing, Cloud Adoption, Cloud economics, security, compliance, Cloud storage (object), Data Protection, Archival, Cloud tiering**

I. Introduction

The cloud computing technology although new but it has shaken the world upside down. It made its emergence as a storm in the IT industry. In very less time, it became popular and people started talking about this technology.

A. History of Cloud Computing

If we look into the history of Cloud Computing, it all started with the emergence of Internet as its pillar. The remarkable development which Cloud Computing has made in its history for the first time is with the arrival of Salesforce.com in the year 1999, delivering enterprise applications via simple website. Amazon Web Services (AWS) marks the official birth of Cloud Computing on August 2006. To add to this, Amazon is still the leader in Public Cloud Market. Cloud Computing became a popular name after the spring of 2006 and the fall of that same year when Amazon came up with its compute service named Amazon EC2 and storage service named Amazon S3. Post the year, many big companies tried to hold their share of billions of dollars in the Cloud Market, Google, Microsoft IBM, HP were some to name. The year 2010 came with a new competitor named Azure which is one of the biggest competitors of AWS in Public Cloud Market.

The different technologies have helped in the development of Cloud Computing which are as follows: -

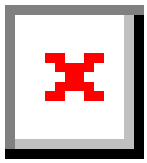
B. Definition of Cloud Computing

Defining the term Cloud Computing is indeed impossible as we have several of Cloud vendors offering different kind of activities under the roof of Cloud Computing as per their needs, so a particular and rigid definition cannot be made for Cloud Computing.

"*Cloud Computing can be described as providing reliable, scalable and highly available infrastructure, software and bandwidth on pay as you go basis which can be changed as per the customer needs."*

II. Cloud Offerings

Cloud offerings can be broadly classified into three different categories which are as follows: -



However, as the Cloud is growing and getting mature, the distinction between these three is getting eroded as the services are tightly coupled.

A. *IAAS – Infrastructure as a Service*

An IAAS service provider run and manage server farms which have virtualization software running and allow us to create the virtual machine as per our need. Depending on the Cloud vendor, the operating system of virtual machines can be Linux, Windows or install anything else as per our need. We do not need to manage the hardware and also, we do not have control over the virtualization but other than we can manage everything else right from operating system. EC2 instances offered by AWS and virtual machines offered by Microsoft Azure are examples of IAAS.

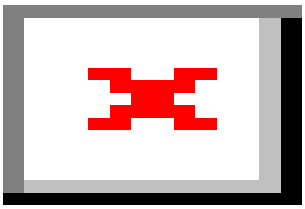
B. *SAAS – Software as a Service*

It is software that is hosted and managed by vendor for the customer. Generally, a single version of application, hosting the software is used for all customers. It can be scaled in and out to multiple instances as per the customer demand. End user doesn't have control over anything and everything like managing the hardware to base operating system is done by SAAS provider. Office 365, Dropbox etc. are example of SAAS.

C. *PAAS – Platform as a Service*

PAAS provides you an application hosting environment where we deploy our application. The end user takes care of the application and ability to run and deploy is managed by Cloud vendor. This frees us from infrastructure management and we can focus on application development only.

The control provided to the end user in different Cloud offering is summarized as below: -



III. WHY TO USE CLOUD

The main reason behind a shift to Cloud for any of the business is speed, scale and economics. Let's discuss about the advantages of using Cloud which is forcing everyone towards Cloud.

Speed: It runs fast which saves our time in deploying, configuring and managing the underlying hardware consisting of compute, storage and networking infrastructure.

Speed: - Developing applications that will run or will be deployed in Cloud is quite fast because we do need to deploy, configure and manage the underlying hardware consisting of compute, storage and networking infrastructure. We can leverage ready-made infrastructure provided by Cloud vendor and there is no headache of procuring the hardware etc. One can easily create replicas of existing environment running on Cloud for different scenarios like Test, Dev or QA.

Scale: It is flexible as per the need of the customer, it can be scaled in or out which reduces the future risks and saves time planning the future risks.

Economics: - One of the biggest factor for adoption of Cloud is cost. Cost of setting up an application on Cloud is comparatively lower as compared to On-Prem. Also, pay-for-what-you-use model also helps organizations to save lot of cost as they can scale in and out as per the demand and pay accordingly.

Flexibility: - Different kind of applications have different requirements in terms of infrastructure and Cloud vendors provide variety of options to the end customer.

Global Operations: - Cloud vendors have global presence i.e. their data centers are spread across the world. So, we can choose a region where we want to deploy the environment as per our needs like if there is specific need w.r.t. compliance that data should not leave a certain physical region.

IV. RESISTING THE CLOUD

If Cloud is really so good and there are so many advantages, why doesn't every business adopt it. The possible for reasons for this are as follows: -

Businesses that have already made huge investments to setup IT infrastructure will prefer to have those investments return them something rather than going for Cloud.

Some of the Enterprises have a complex IT environment which makes it difficult for them to migrate to Cloud, or integrate Cloud Computing into their existing infrastructure.

As on Cloud we have less transparency as compared to in-house IT Infrastructure in terms of hardware used. The reliability and faith with hosting provider is an issue for many companies.

Organization policies can also be one reason that resist migrating their applications on to the Cloud.

Organizations that needs to follow certain compliance and have strict regulation in the areas of security sometimes have legal or regulatory requirements which doesn't allow them to adapt Cloud Computing.

Are You Ready for Cloud?

There is a big question that is *Cloud* good for every business and any application can be taken to *Cloud*. Every business is not ready to integrate and leverage different types of *Cloud Computing* services available. There should be a proper analysis of application and environment before we can adopt

Cloud. Different *Cloud* have their own operational checklist to analyze the *Cloud* readiness of an application and use cases.

But there are certain general Cloud procurement consideration in terms of management which are as follows: -

Billing and Account Governance

Security and Access Management

Monitoring and Incident Management

Configuration and Change Management

Release and Deployment Management

Asset Management

Application High availability

Application DR

There are different perspectives for Cloud Adoption Framework which is as follows: -

Business Perspective – it represents the areas that one focus to ensure that technology is utilized in an efficient way.

Platform Perspective – it represents the areas that one focus to ensure environment is architected and designed in a way to achieve the expected levels of functionality

Maturity Perspective – it represents the areas that one focus to ensure that correct initial assessment of the current state is done and desired target state is defined.

People Perspective – it represents the areas that one focus to ensure organizational structures and competency is there to implement, operate and manage Cloud-based environment.

Process Perspective – it represents the areas that one focus to ensure processes are in place to implement, operate and manage Cloud-based environment.

Operations Perspective – it represents the areas that one focus to ensure the Cloud environment can be efficiently operated and managed.

Security Perspective – it represents the areas that one focus on implementing security within the Cloud environment.

V. STORAGE ON CLOUD

One of the major roadblocks associated with Cloud is Data Storage and Backup because data is the most precious asset for any Enterprise and while leveraging public Cloud, data is going to be at Cloud vendor location and we don't have the visibility of location, security and access. Let's discuss how Cloud storage can be efficiently used.

A. How we can efficiently use storage on Cloud

Cloud Storage (Object Storage) is highly reliable object storage which can be very useful in storing cold data. With reliability, this is very cost efficient as well and can help organizations to store huge amount of data with least burden on their IT budget.

Cloud storage is revolutionary when we use it as target for storing backup data (creating and storing another copy of production data) or archival of data (removing less frequently accessed data from costly storage volumes to blob storage)

B. Backup (Data Protection) to Cloud storage

Backup to Cloud is gaining a lot of attention now a day and everyone is curious to know if we can move our backups to Cloud storage as target.

Traditionally tapes were used as target for backups. Then disk was selected as preferred medium for faster issues and to get rid of issues with tapes and tape management. Deduplication was added to disk based backups to make sure we get some control on disk storage which was expanding exponentially with backup data. Now, companies are exploring Cloud storage as target for backups due to its cost benefits (3-4 cents/GB) and reliability (much better than tapes).

First of all, we need to understand and identify the type of backups which can be moved to Cloud. It is very important to understand that Cloud cannot be target for all of your backup data. You still need either tapes or disks to store the backups which are frequently accessed and should be readily available for restore. Typically, all daily and weekly backups fall in this category.

Many companies have long term retention requirement which can vary from 1 year to infinite, based on compliance and legal requirements. These backups should be selected while we try to move backups to Cloud (typically monthly and yearly backups).

Cloud Target is for backup data which is rarely accessed and chances of getting restore is minimum due to following facts:

Data out charges will apply for any restore done from Cloud storage

Network bandwidth needs to be planned and restore might take a long time (longer than disk) depending on the bandwidth.

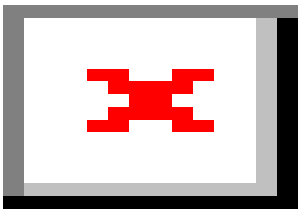
Frequent restore requests will increase the cost of backup solution and can also cause issues while customers are trying to meet their SLA's for critical applications (limited Network Bandwidth).

Having said that, there are lot of benefits to move backups to Cloud. Cloud storage is a cost effective and reliable solution for long term retention backups when compared with tapes.

Backup of On-Premise data on Cloud storage is helpful if we have infrequent access to data and long term retention. So, Monthly/Yearly backups are ideal for the same and cost of

selecting Cloud over tapes is cheaper or competitive in such cases.
Also, with Cloud we get more reliability in case we need to restore the data which was backed up 5-7 years back (Tape generation keeps on changing and even if we backup for 7 years, after 7 years getting the restore done is a complicated process).

Approach for Data Protection to Cloud can be summarized in below figure:



C. Archival to Cloud

Another interesting use case of utilizing cost efficient and reliable Cloud storage is Archival of data. For this, first step is to do e-discovery of data in your environment and based on that do data classification based on below parameters:

Critical data
Important data
Standard data

- Frequently accessed
- Application data being accessed continuously by application.
- Cold data - not being accessed or modified from a long time.

For Archival to Cloud, cold data is ideal and once we come to a conclusion that this is not being accessed from a particular time period (defined by end users), we can achieve it to Cloud storage (blob) with pointers on end user side.

Advantages:
Archival to Cloud will assist in reducing the backup volume and Storage footprints for customers having infrequent access to file data.
We can decide a time limit and if the data is not access within that time, it will be archived to Cloud with option of having a stub/no stub. Retention will also be defined to data residing on Cloud, after which it will expire from Cloud target as well.

Disadvantages:
Getting the Archived file back from Cloud is similar to downloading the file from internet and depends on the network bandwidth between customers Datacenters and Cloud provider. We need to size the same and make sure that customers' expectations are in sync with the same.

D. Hybrid Storage with Cloud:

With data growing exponentially, IT is also moving towards utilizing Cloud storage as an extended tier to reduce their storage footprints in Datacenters.

Advantages:
IT product companies are now providing a solution, where we have the flexibility to tier our storage and it has intelligence of sending the data to Cloud with benefits of deduplication, compression and wan optimization.

This will be helpful in scenarios where customer is looking for scalable storage which can scale to Cloud as well and provide mix of On-premise and Cloud storage. This will be provided as iSCSI storage and will be mapped to servers directly.

Disadvantages:
We can't compare this with SAN or high performance Storage. This is for medium workloads and its performance will largely depend on the local SSD drives and SATA disks which is limited per device.
Few points to take care while doing data classification for Tiered Storage:

Non-mission critical data
File shares (older files, not accessed frequently)
Unstructured data
Data which is not accessed frequently – Very Important.

VI. CONCLUSION

Cloud Computing is a marketing gimmick and analysis needed to be done on various parameters before you actually migrate applications to cloud. You should analyze your needs and walk your application through Cloud Readiness check. Also you must think in terms of commercials, security, network, data protection and storage. Cloud might not be the right fit for every business - think before you act!

VII. REFERENCES

[1]. https://en.wikipedia.org/wiki/Object_storage
[2]. https://en.wikipedia.org/wiki/Cloud_Computing
[3]. https://aws.amazon.com/
[4]. https://azure.microsoft.com/
[5]. https://aws.amazon.com/whitepapers

Cloud Computing – Think before you Act

Analyzing Cloud Computing on key Parameters

Amit Kumar
AWS Certified
IINTM
New Delhi, India
amkr95@gmail.com

Tomesh
IINTM
New Delhi, India

Achint Sharma
IINTM
New Delhi, India

Abstract— In past few years, Cloud Computing has become a big word in IT industry and everyone is talking about it. But there are more questions associated with it rather than the answers, is it good for everyone and if it is then why everyone is not adopting it with open hands. Cloud Computing is like a dish, about which everyone talks about but no one seems to be in a hurry to taste it. On the basis of accessibility and resource location Cloud can be categorized as Public Cloud, Private Cloud and Hybrid Cloud. In this paper, mostly we are going to focus on Public Cloud. Also, we will discuss about Cloud Computing in terms of adoption feasibility, economics, ease of use, security and various advantages & disadvantages.

Keywords— Cloud Computing, Cloud Adoption, Cloud economics, security, compliance, Cloud storage (object), Data Protection, Archival, Cloud tiering

I. INTRODUCTION

Cloud Computing has come as a storm in the IT world. It seems to be most talked about technology in the past few years. Though Cloud Computing is quite a new technology but it has already impacted every industry in the world and has shaken the world upside down.

A. History of Cloud Computing

The story of Cloud Computing starts with the emergence of Internet in the world as it is basic pillar for Cloud Computing. The first major development in the Cloud Computing history was arrival of Salesforce.com in 1999, which delivers enterprise applications via a simple website. Salesforce.com started the era where software is served as a service over the internet. But officially the birth of Cloud Computing can be considered as August, 2006 when Amazon launched Amazon Web Services (AWS), which is still the leader in Public Cloud Market. At that time, Amazon's target was developers and small companies who don't want to get into headache of maintaining the IT infrastructure. At that time nobody spokes about the term Cloud Computing and this concept was more like renting out the IT infrastructure. AWS came up with first service which is storage service name Amazon S3 in the spring of 2006 and compute service named Amazon EC2 in the fall of that year. After that the development of AWS till today is a public history of which everyone talks about. After that many other small and big players came into the market to capture their share of billions of dollars of Cloud Market, some of which are Microsoft, IBM, Google, HP etc. Microsoft Azure was released on February, 2010 and is one of the biggest competitors of AWS in Public Cloud Market.

The different technologies have helped in the development of Cloud Computing which are as follows:-

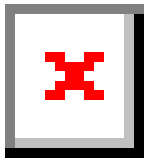
B. Definition of Cloud Computing

Defining the term Cloud Computing is almost an impossible task as different Cloud vendors provide different kind of services under the hood of Cloud Computing, also the users are using different services as per their needs.

"*Cloud Computing can be described as providing reliable, scalable and highly available infrastructure, software and bandwidth on pay as you go basis which can be changed as per the customer needs."*

II. CLOUD OFFERINGS

Cloud offerings can be broadly classified into three different categories which are as follows:-



However as the Cloud is growing and getting mature, the distinction between these three is getting eroded as the services are tightly coupled.

A. *IAAS – Infrastructure as a Service*

An IAAS service provider run and manage server farms which have virtualization software running and allow us to create the virtual machine as per our need. Depending on the Cloud vendor, the operating system of virtual machines can be Linux, Windows or install anything else as per our need. We do not need to manage the hardware and also we do not have control over the virtualization but other than we can manage everything else right from operating system. EC2 instances offered by AWS and virtual machines offered by Microsoft Azure are examples of IAAS.

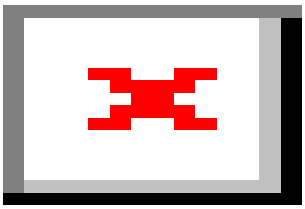
B. *SAAS – Software as a Service*

It is software that is hosted and managed by vendor for the customer. Generally a single version of application, hosting the software is used for all customers. It can be scaled in and out to multiple instances as per the customer demand. End user doesn't have control over anything and everything like managing the hardware to base operating system is done by SAAS provider. Office 365, Dropbox etc. are example of SAAS.

C. *PAAS – Platform as a Service*

PAAS provides you an application hosting environment where we deploy our application. The end user takes care of the application and ability to run and deploy is managed by Cloud vendor. This frees us from infrastructure management and we can focus on application development only.

The control provided to the end user in different Cloud offering is summarized as below:-



III. WHY TO USE CLOUD

The main reason behind a shift to Cloud for any of the business is speed, scale and economics. Let's discuss about the advantages of using Cloud which is forcing everyone towards Cloud.

Speed: - Developing applications that will run or will be deployed in Cloud is quite fast because we do need to deploy, configure and manage the underlying hardware consisting of compute, storage and networking infrastructure. We can leverage ready-made infrastructure provided by Cloud vendor and there is no headache of procuring the hardware etc. One can easily create replicas of existing environment running on Cloud for different scenarios like Test, Dev or QA.

Scale: - Cloud applications also scale in and out very quickly as per the needs. There is no need to predict the future growth and procure the hardware accordingly right from the first day.

Economics: - One of the biggest factor for adoption of Cloud is cost. Cost of setting up an application on Cloud is comparatively lower as compared to On-Prem. Also pay-for-what-you-use model also helps organizations to save lot of cost as they can scale in and out as per the demand and pay accordingly.

Flexibility: - Different kind of applications have different requirements in terms of infrastructure and Cloud vendors provide variety of options to the end customer.

Global Operations: - Cloud vendors have global presence i.e. their data centers are spread across the world. So we can choose a region where we want to deploy the environment as per our needs like if there is specific need w.r.t. compliance that data should not leave a certain physical region.

IV. RESISTING THE CLOUD

If Cloud is really so good and there are so many advantages, why doesn't every business adopt it. The possible for reasons for this are as follows:-

Businesses that have already made huge investments to setup IT infrastructure will prefer to have those investments return them something rather than going for Cloud.

Some of the Enterprises have a complex IT environment which makes it difficult for them to migrate to Cloud, or integrate Cloud Computing into their existing infrastructure.

As on Cloud we have less transparency as compared to in-house IT Infrastructure in terms of hardware used. The reliability and faith with hosting provider is an issue for many companies.

Organization policies can also be one reason that resist migrating their applications on to the Cloud.

Organizations that needs to follow certain compliance and have strict regulation in the areas of security sometimes have legal or regulatory requirements which doesn't allow them to adapt Cloud Computing.

Are You Ready For Cloud:

There is a big question that is *Cloud* good for every business and any application can be taken to *Cloud*. Every business is not ready to integrate and leverage different types of *Cloud Computing* services available. There should be a proper analysis of application and environment before we can adopt

Cloud. Different *Cloud* have their own operational checklist to analyze the *Cloud* readiness of an application and use cases.

But there are certain general Cloud procurement consideration in terms of management which are as follows:-

Billing and Account Governance

Security and Access Management

Monitoring and Incident Management

Configuration and Change Management

Release and Deployment Management

Asset Management

Application High availability

Application DR

There are different perspectives for Cloud Adoption Framework which is as follows:-

Business Perspective – it represents the areas that one focus to ensure that technology is utilized in an efficient way.

Platform Perspective – it represents the areas that one focus to ensure environment is architected and designed in a way to achieve the expected levels of functionality

Maturity Perspective – it represents the areas that one focus to ensure that correct initial assessment of the current state is done and desired target state is defined.

People Perspective – it represents the areas that one focus to ensure organizational structures and competency is there to implement, operate and manage Cloud-based environment.

Process Perspective – it represents the areas that one focus to ensure processes are in place to implement, operate and manage Cloud-based environment.

Operations Perspective – it represents the areas that one focus to ensure the Cloud environment can be efficiently operated and managed.

Security Perspective – it represents the areas that one focus on implementing security within the Cloud environment.

V. STORAGE ON CLOUD

One of the major roadblocks associated with Cloud is Data Storage and Backup because data is the most precious asset for any Enterprise and while leveraging public Cloud, data is going to be at Cloud vendor location and we don't have the visibility of location, security and access. Let's discuss how Cloud storage can be efficiently used.

A. How we can efficiently use storage on Cloud

Cloud Storage (Object Storage) is highly reliable object storage which can be very useful in storing cold data. With reliability, this is very cost efficient as well and can help organizations to store huge amount of data with least burden on their IT budget.

Cloud storage is revolutionary when we use it as target for storing backup data (creating and storing another copy of production data) or archival of data (removing less frequently accessed data from costly storage volumes to blob storage)

B. Backup (Data Protection) to Cloud storage

Backup to Cloud is gaining a lot of attention now a days and everyone is curious to know if we can move our backups to Cloud storage as target.

Traditionally tapes were used as target for backups. Then disk was selected as preferred medium for faster issues and to get rid of issues with tapes and tape management. Deduplication was added to disk based backups to make sure we get some control on disk storage which was expanding exponentially with backup data .Now , companies are exploring Cloud storage as target for backups due to its cost benefits (3-4 cents/GB) and reliability (much better than tapes).

First of all we need to understand and identify the type of backups which can be moved to Cloud. It is very important to understand that Cloud cannot be target for all of your backup data. You still need either tapes or disks to store the backups which are frequently accessed and should be readily available for restore. Typically all daily and weekly backups falls in this category.

Many companies have long term retention requirement which can vary from 1 year to infinite, based on compliance and legal requirements. These backups should be selected while we try to move backups to Cloud (typically monthly and yearly backups).

Cloud Target is for backup data which is rarely accessed and chances of getting restore is minimum due to following facts:

Data out charges will apply for any restore done from Cloud storage

Network bandwidth needs to be planned and restore might take a long time (longer than disk) depending on the bandwidth.

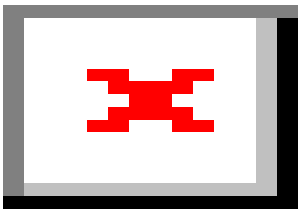
Frequent restore requests will increase the cost of backup solution and can also cause issues while customers are trying to meet their SLA's for critical applications (limited Network Bandwidth).

Having said that, there are lot of benefits to move backups to Cloud. Cloud storage is a cost effective and reliable solution for long term retention backups when compared with tapes.

Backup of On-Premise data on Cloud storage is helpful if we have infrequent access to data and long term retention. So, Monthly/Yearly backups are ideal for the same and cost of

selecting Cloud over tapes is cheaper or competitive in such cases.
Also , with Cloud we get more reliability in case we need to restore the data which was backed up 5-7 years back (Tape generation keeps on changing and even if we backup for 7 years , after 7 years getting the restore done is a complicated process).

Approach for Data Protection to Cloud can be summarized in below figure:



C. Archival to Cloud

Another interesting use case of utilizing cost efficient and reliable Cloud storage is Archival of data. For this, first step is to do e-discovery of data in your environment and based on that do data classification based on below parameters:

Critical data
Important data
Standard data

- Frequently accessed
- Application data being accessed continuously by application.
- Cold data - not being accessed or modified from a long time.

For Archival to Cloud, cold data is ideal and once we come to a conclusion that this is not being accessed from a particular time period (defined by end users), we can achieve it to Cloud storage (blob) with pointers on end user side.

Advantages:
Archival to Cloud will assist in reducing the backup volume and Storage footprints for customers having infrequent access to file data.
We can decide a time limit and if the data is not access within that time, it will be archived to Cloud with option of having a stub/no stub. Retention will also be defined to data residing on Cloud, after which it will expire from Cloud target as well.

Disadvantages:
Getting the Archived file back from Cloud is similar to downloading the file from internet and depends on the network bandwidth between customers Datacenters and Cloud provider. We need to size the same and make sure that customers' expectations are in sync with the same.

D. Hybrid Storage with Cloud:

With data growing exponentially, IT is also moving towards utilizing Cloud storage as an extended tier to reduce their storage footprints in Datacenters.

Advantages:
IT product companies are now providing a solution, where we have the flexibility to tier our storage and it has intelligence of sending the data to Cloud with benefits of deduplication, compression and wan optimization.

This will be helpful in scenarios where customer is looking for scalable storage which can scale to Cloud as well and provide mix of On-premise and Cloud storage. This will be provided as iSCSI storage and will be mapped to servers directly.

Disadvantages:
We can't compare this with SAN or high performance Storage. This is for medium workloads and its performance will largely depend on the local SSD drives and SATA disks which is limited per device.
Few points to take care while doing data classification for Tiered Storage:

Non-mission critical data
File shares (older files , not accessed frequently)
Unstructured data
Data which is not accessed frequently – Very Important.

VI. CONCLUSION

Cloud Computing is a marketing gimmick and analysis needed to be done on various parameters before you actually migrate applications to cloud. You should analyze your needs and walk your application through Cloud Readiness check. Also you must think in terms of commercials, security, network, data protection and storage. Cloud might not be the right fit for every business - think before you act!

VII. REFERENCES

[1]. Cloud Computing: Concepts, Technology & Architecture – By Ricardo Puttini, Thomas Erl, and Zaigham Mahmood
[2]. Cloud Computing For Dummies- By Fern Halper, Judith Hurwitz, Marcia Kaufman, and Robin Bloor
[3]. https://aws.amazon.com/
[4]. https://azure.microsoft.com/
[5]. https://aws.amazon.com/whitepapers

DIGITAL SECURITY

Submited to:
Mr.Ajay Phogat

Submited By:
ANCHAL KUNDRA-05590302014
GAGAN CHAUHAN-05690302014
RAVEENA RATHORE-05790302014

Digital Security

Raveena Rathore
Department of Information & Technology,
Institute of Innovation in Technology & Management,
G.G.S.I.P.U, New Delhi
Email ID: raveenan6500s@gmail.com

Anchal Kundra
Department of Information & Technology,
Institute of Innovation in Technology & Management,
G.G.S.I.P.U, New Delhi
Email ID: anchalkundra12@gmail.com

ABSTRACT

Digital security is the protection of your digital identity- the network or Internet equivalent of your physical identity. Digital security includes the tools you use to secure your identity, assets and technology in the online and mobile world. Digital Security is the precautionary protection of your online identity. This means not providing your information online where others can see it. This can include your full name, address, passwords, bank information, etc. the tools used in digital security include anti-virus softwares, Web services, biometrics and secure personal devices you carry with you every day. Secure personal devices such as smart card-based USB token, the SIM card in your cell phone or e-passport are digital devices because they give you freedom to communicate, travel, shop, bank and work using your digital identity in a way that is convenient, enjoyable and secure.

Digital Security is important because it allows people to use social media and online baking and protects them from risks such as identity theft and fraud. If the steps mentioned earlier are followed, then your digital security is strong and will protect your information. It is important to also be aware that your photos should be on private because they can be used when stealing your identity. It is also important not to open any of your important accounts such as your inline bank accounts in public because it might results in the people around you looking at your information and obtaining your important and private information.

1. INTRODUCTION

Digital security is about regulating access to digital assets, which can be information or services. Good digital security begins with security requirement engineering, *i.e.* by identifying the actors, their assets and interests, and their authorization levels (who is allowed to do what). This should lead to suitable mechanisms in the form of security policies, protocols, and rules to achieve the "right" level of security in a given context, and finally providing secure software and hardware that realize these mechanisms. These security mechanisms may not only be used for restricting users to behave properly, but also for monitoring and detecting improper behaviour.

One topic of research here is identity-centric security, which focuses on identity management. This includes the investigation of the policies and protocols for identity management, the mechanisms such as smart cards, RFID tags, or biometrics that can be used for this, and their impact on privacy and anonymity. Text and data mining for finding structure or patterns in large amounts of data can also be used to our advantage, to detect anomalies or produce warnings, for instance in a cybercrime context.

Another topic of research is software security, which looks at the role that software plays on the one hand in providing security, and on the other hand as a source of security vulnerabilities. The focus is on ways to ensure the correct implementation of security functionality and the absence of security vulnerabilities, by formal specification of security properties of code, and checking these by means of verification, typing, (penetration) testing, or code inspection.

A broader topic of research is the formulation and formalisations of security policies and security rules, and methods for risk management and risk assessment. There is a clear overlap between the theme of digital security and model based system development, in that specification and analysis of security relies on models, which are also a basis for verification and testing. There is also an overlap with intelligent systems, because intelligent systems can for instance be used to

detect anomalies and trigger alerts, and to support the verification process.

2. METHODOLOGY

The research was inspired by a report from Harvard University's Berkman Center for Internet and Society called "Online Security in the Middle East and North Africa: A Survey of Perceptions, Knowledge, Practice."[4] As researchers were interested in comparisons between regions, some of the questions asked in the Pakistan survey were similar to those in the Berkman report.

2.1. Guiding Principles

This research was intended to be used to develop training workshops for journalists and bloggers to address the issues they face in their daily lives regarding online security. This study could also lay the groundwork for more in-depth research on this topic.

2.1.1. SAMPLE SELECTION

Potential respondents were selected using convenience sampling, on the basis of their importance in the media world and the blogosphere. Care was taken to ensure gender, regional diversity, and national scope among participants.

Contact was made through telephone, email and various sources within the journalist community. A total of 52 people (65% of those initially contacted) completed questionnaires. Seventy percent of the respondents were working journalists and the remaining 30% identified themselves as bloggers.

2.1.2. SURVEY QUESTION DESIGN

In designing the survey questions, Internews ensured that the questions asked were relevant to both journalists and bloggers. The same questions were asked from both because the fields are interlinked: most Pakistani journalists maintain blogs, and most bloggers write posts that are at times journalistic.

2.1.3. LANGUAGE

Instead of focusing only on English-language media, journalists from Urdu- and regional language publications were included as well. Respondents had the option of selecting the language in which they could most comfortably interact. Researchers conducted the interviews in English, Urdu, and Pashto.

2.1.4. CROSS-COUNTRY SAMPLING

The researchers conducted cross-country sampling for this research, interviewing journalists and bloggers from around Pakistan to get a diversity of perspectives. Respondents' regions broke down as follows:

Area of Residence	Number of respondents
Punjab	20
Sindh	12
Khyber Pakhtunkhwa	12
Islamabad	8

3. USE OF DIGITAL TECHNOLOGY

When researchers asked participating journalists and bloggers about the use of technology in their work, it was apparent respondents are very wired. Nearly all use desktop or laptop computers and most use mobile phones and social networking websites in their day-to-day work. Over 90% use the Internet in the course of their professional duties, and most of those use it for story research.

3.1. Usage of Internet for Research

Nearly 81% of respondents use the Internet for story research, with over half reporting heavy use, as below.

Using the Internet for Story Research	%
No use	0.0
Low use	17.3
Moderate use	26.9
Heavy use	53.8

3.2. Social Media Usage

In researching, distributing or writing a story, respondents reported that Facebook, YouTube and Twitter were their three most used social media platforms. Fewer than 5% chose Google+, Flickr or Orkut. Facebook is without a doubt the most popular social networking website, with 75% of

respondents using it to a moderate or heavy extent, followed by YouTube at 63.5% and then Twitter with 51.9%.

Social Media	Facebook %	YouTube %	Twitter %
No use	13.5	36.5	48.1
Low use	11.5	0.0	0.0
Moderate use	32.7	50	19.2
Heavy use	42.3	13.5	32.7

3.3. Email Provider Preference

When asked which email service they use for their work communications, 76.9% of respondents reported using Gmail. This was an encouraging result, because Gmail is relatively secure compared to other email service providers like Hotmail and Yahoo. Very few respondents use their company's private email services. According to some respondents, they prefer using Gmail even though their company provides them with an exclusive email address.

Email service	%
Gmail	76.9
Hotmail	13.5
Company email	7.7
Yahoo	5.8

Survey participants were asked what features are most important to them in selecting an email service. The results clearly explain the preference for Gmail. The largest group of respondents (49.2%) replied that 'storage space' is the most important feature for them in an email service. Gmail offers around 7.5 gigabytes (GB) of free space. The second most important feature for respondents was 'ease of use.'

Only 18.6% of those polled considered 'security' to be the most important feature for them to have in an email service. While this figure is relatively high compared to other countries, the authors of this report are concerned that journalists and bloggers, professions that face special security threats, do not take their digital security as seriously as they should.

selecting an email service	%
Security	18.6
Storage space	49.2
Ease of use	30.5
Any other	0.0

3.4. Awareness of Secure Email Features

The majority of respondents (60%) were unaware of the existence of secure email features, such as point-to-point encryption, where emails sent to and received from the respondent's computer to email providers' servers are encrypted using the secure sockets layer (SSL) protocol. Google's email client, Gmail, provides this feature; however the survey responses suggest that a majority of users are not fully aware of this.

3.5 Blogging and Micro-blogging Provider Preference

Most respondents rated either 'ease of use' (39.1%) or 'ability to customize' (34.8%) as the most important feature in selecting a blogging or micro-blogging service such as Twitter. 'Popularity' was another top consideration. Once again, as with email service selection, 'security/privacy' was not much of a factor: only 6.5% respondents said that they make the security features of a particular blogging or micro-blogging portal their first priority in making their decision.

ETHICAL HACKING

Aayushi , Meenu , Samish

Institute of Innovation in Technology & Management

New Delhi

aayushisingh221@gmail.com

meenumittal497@gmail.com

samishdon@gmail.com

ABSTRACT

The explosive growth of the Internet has brought many good things such as E-commerce-banking, E-mail, Cloud Computing, but there is also a Dark side such as Hacking, Backdoors etc. Hacking is the first big problem faced by Governments, companies, and private citizens around the world , Hacking includes reading others e-mail, steal their credit card number from an on-line shopping site, secretly transmitting secrets to the open Internet. An Ethical Hacker can help the people who are suffered by this Hackings.

Kevin Mitnick, often incorrectly called by many God of hackers, broke into the computer systems of the World's top technology and telecommunications companies Nokia, Fujitsu, Motorola, and Sun Micro systems. He was arrested by the FBI in 1995, but later released on parole in 2000. He never termed his activity hacking, instead he called it social engineering.

Successful ethical hackers possess a variety of skills. First and foremost, they must be completely trustworthy. Ethical hacking is not just necessary ; it is inevitable.

Keywords:

1. Ethical hackers- a person who hacks into a computer network in order to test or evaluate its security, rather than with malicious or criminal intent.
2. White hats- a person who hacks into a computer network in order to test or evaluate its security systems.
3. Malicious hackers- (black hat) describes any individual who illegally breaks into computer systems to damage or steal information.
4. Grey hats- a computer hacker or computer security expert who may sometimes violate laws or typical ethical standards, but does not have the malicious intent typical of a black hat hacker.
5. Cyber security- the body of technologies, processes and practices designed to protect networks, computers, programs and data from attack, damage or unauthorized access.

I. INTRODUCTION

Security is the major fact in today's era where internet use is very vast and fast growing. Every organization has issues related to security about their sensitive and confidential data. This is only because of hacking, Hacking is done by a person who has wrong intensions. Basically there are two types of hackers, one who has rights of securing data while using hacking techniques and the other who uses his knowledge to break security to harm the organization. These hackers are categorized into two categories :-

1. **Ethical Hackers**
2. **Malicious Hackers**

Hacking is a process of controlling the system of an organization without the knowledge of the organization members. In contrast it is called breaking the security to steal the sensitive and confidential information such as credit card numbers, telephone numbers, home addresses, bank account numbers etc that are available on network. This illustrates that security is a discipline which protects the confidentiality, integrity & availability of resources. It refers this era as a "Security Era" not

because we are very much concerned about security but due to the maximum need of security . It also explains that the explosive growth of internet has brought many good things such as electronic commerce, easy access to vast stores of reference material, collaborative computing, email and new avenues of advertising and information distribution etc. but there is also a dark side such as criminal hackers. The government, companies and private citizens around the world are anxious to be a part of this revolution, but they are very much afraid that some hackers will break into their Web Server and replaces their information with pornography, read their e-mail, steal their credit card number from an on-line shopping site, or implant software that will secretly transmit their organization information to the open internet. Cyber Security is the most talked about topic and the most concerned area in today's online world.

Survey on Cyber Security

I noticed on a report from government website that is actually "Internet Crime Current Report". The Internet Crime Complaint Center (IC3) is a partnership between the Federal Bureau of Investigation (FBI) and the National White Collar Crime Center (NWC3).

Figure 1: Complaint Graph of IC3

The graph (Fig. 1) shows that till year 2009 numbers of complaint were increased at an exponential rate. But after 2009 it is little down in 2010 but increased in 2011 and is almost the same in the remaining years.

This shows that how much important the security is in this era of computer world.

II. ABOUT HACKING

Eric Raymond, compiler of "**The New Hacker's Dictionary**", defines a hacker as a clever programmer. A "good hack" is a clever solution to a programming problem and "hacking" is the act of doing it. Raymond lists five possible characteristics that qualify one as a hacker, which we paraphrase here:

- A person who enjoys learning details of a programming language or system
- A person who enjoys actually doing the programming rather than just theorizing about it
- A person capable of appreciating someone else's hacking
- A person who picks up programming quickly
- A person who is an expert at a particular programming language or system

TYPES OF HACKERS:

Hackers can be broadly classified on the basis of why they are hacking system or why the are indulging hacking. There are mainly three types of hacker on this basis:-

Figure 2: Types of Hackers

- **Black-Hat Hacker** - A black hat hackers or crackers are individuals with extraordinary computing skills, resorting to malicious or destructive activities. That is black hat hackers use their knowledge and skill for their own personal gains probably by hurting others.
- **White-Hat Hacker-** White hat hackers are those individuals professing hacker skills and using them for defensive purposes. This means that the white hat

hackers use their knowledge and skill for the good of others and for the common good.

● **Grey-Hat Hackers**- These are individuals who work both offensively and defensively at various times. We cannot predict their behaviour. Sometimes they use their skills for the common good while in some other times he uses them for their personal gains.

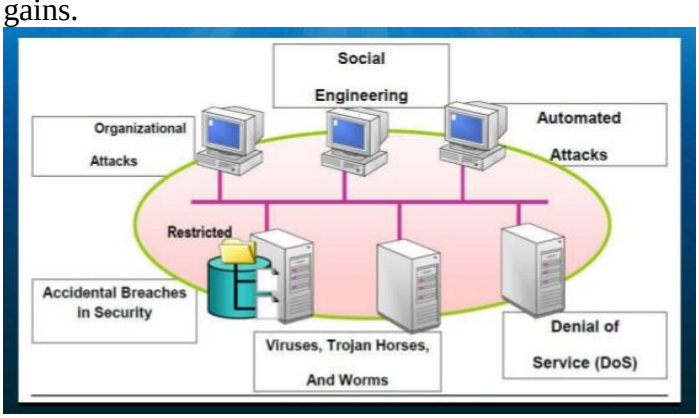


Figure3: Different kinds of System attacks

III . ETHICAL HACKING

1) **Ethical hacking** – defined as "a methodology adopted by ethical hackers to discover the vulnerabilities existing in information systems' operating environments."
2) With the growth of the Internet, computer security has become a major concern for businesses and governments.
3) In their search for a way to approach the problem, organizations came to realize that one of the best ways to evaluate the intruder threat to their interests would be to have independent computer security professionals attempt to break into their computer systems.

What do an Ethical Hacker do?

An ethical hacker is a person doing ethical hacking that is he is a security personal who tries to penetrate in to a network to find if there is some vulnerability in the system. He will first find out what an intruder can see or what others can see. Finding these an ethical hacker will try to get into the system with that information in whatever method he can. If he succeeds in penetrating into the system then he will report to the company with a detailed report about the particular vulnerability exploiting which he got in to the system. He may also sometimes make patches for that particular vulnerability or he may suggest some methods to prevent the vulnerability.

Required skills of an ethical hacker:

1) Microsoft: skills in operation, configuration and management.
2) Linux: knowledge of Linux/Unix; security setting, configuration, and services.
3) Firewalls: configurations, and operation of intrusion detection systems.
4) Routers: knowledge of routers, routing protocols, and access control lists
5) Network Protocols: TCP/IP; how they function and can be manipulated.
6) Project Management: leading, planning, organizing, and controlling a penetration testing team.

ETHICAL HACKING COMMANDMENTS:

Every ethical hacker must abide by a few basic commandments. If not, bad things can happen. The commandments are as follows:

Working ethically:

The word ethical in this context can be defined as working with high professional morals and principles. Everything you do as an ethical hacker must be aboveboard and must support the company's goals. No hidden agendas are allowed! Trustworthiness is the ultimate tenet. The misuse of information is absolutely forbidden.

Respecting privacy:

Treat the information gathered with the utmost respect. All information you obtain during your testing — from Web-application log files to clear-text passwords — must be kept private. If you sense that someone should know there's a problem, consider sharing that information with the appropriate manager.

Not crashing your systems:

One of the biggest mistakes hackers try to hack their own systems is inadvertently crashing their systems. The main reason for this is poor planning. These testers have not read the documentation or misunderstand the usage and power of the security tools and techniques.

METHODOLOGY OF HACKING:

As described above there are mainly five steps in hacking. But it is not the end of the process. The actual hacking will be a circular one. Once the hacker completed the five steps then the hacker will start reconnaissance in that stage and the preceding stages to get in to the next level.

Figure4: Ethical Hacking Phases

1. Reconnaissance: It refers to gather as more information as we can about target in prior to perform an attack. The information is gathered regarding the target without knowledge of targeted company (or Individual).It could be done simply by searching information of target on internet or bribing an employee of the targeted company who would reveal and provide useful information to the hacker. This process is also called as "information gathering". It can be Active or Passive.

2. Scanning: It refers to scan for all the open as well as closed ports and even for the known vulnerabilities on the target machine.

3. Gaining access: It can be gained at OS level, system level or even network level. From normal access hacker can even proceed with privilege escalation. It often includes password cracking, buffer overflows, DOS attack etc.

4. Maintaining Access: It is where hacker strives to retain it control over target with backdoors, root kits or Trojans. Compromised machines can even be used as Bots and Zombies for further attacks.

5. Clearing Logs: It is also known as Daisy Chaining. To avoid being exposed or caught, a good hacker will leave no impressions of his presence. So he attempts to overwrite the system and application logs.

MAIN BENEFITS OF ETHICAL HACKING:-

As ethical hacking plays an important role in this security era where the network users are increasing frequently and also the hackers who are taking advantages of network while sitting at their home. The ethical hacking has following advantages .

1. The fight against terrorism and security issues.
2. Preventing malicious hackers to gain access of crucial data.
3. Ethical hackers believe one can best protect systems by probing them while causing no damage and subsequently fixing the vulnerabilities found.
4. Ethical hackers use their knowledge as risk management techniques.

IV. FUTURE ENHANCEMENTS

- ✓ As it an evolving branch the scope of enhancement in technology is immense. No ethical hacker can ensure the system security by using the same technique repeatedly. He would have to improve, develop and explore new avenues repeatedly.

- ✓ More enhanced software's should be used for optimum protection. Tools used, need to be updated regularly and more efficient ones need to be developed.

V. CONCLUSION

As the use of internet increases, everyone becomes dependent over it and saves their crucial and important data over the internet. Basically this is an invitation to the "crackers" to gain access of information. Thus security is the major problem for the organization. This illustrates the importance of ethical hackers. For this purpose the organization hires ethical hackers who are well knowledge and experienced person. Educate the employees and the users against black hat hacking. Use every possible security measures like Honey pots, Intrusion Detection Systems, Firewalls etc. Every time make our password strong by making it harder and longer to be cracked. The final and foremost thing should be to try **ETHICAL HACKING** at regular intervals.

REFERENCES

[1] Need of Ethical Hacking in online World (A research paper by Monika Pangaria & Vivek Shrivastav), www.ijsr.net/archive/v2i4/IJSRON2013859.pdf

[2] Ethical Hacking Techniques with Penetration Testing www.ijcsit.com/docs/Volume%205/vol5issue03/ijcsit20140503161.pdf(by KB Chowdappa)

[3] System Security and Ethical Hacking www.ijreat.org/Papers%202013/Volume1/IJREATV1I1018.pdf

[4] Ethical Hacking by C. C. Palmar, IBM research division

[5] Ethical Hacking by R. Hartley

[6] Compressive Study on EthicalHacking www.ermt.net/docs/papers/Volume_4/1_January2015/V4N1-117.pdf

[7] Internet Crime Complaint Centre link: www.ic3.gov

GI-FI (A New Wireless Technology)

ABSTRACT

The paper presents the review about GI- FI which is a new wireless technology. GI-FI will help to push wireless communications to faster drive. For many years cables were used as a medium of communication or data transfer. Optical fibres played a dominant role for its higher bit rates and faster transmission. But the installation of cables in optical fibre caused a greater difficulty and thus led to wireless access. Initially wireless technology includes Infrared which was a very slow technology further inventions were done to make wireless technology a better for communication and the invention of Bluetooth, Wi-Fi, WIMAX moved wireless communication to a new era. Researches are being made to improve the wireless technology and as a result, moved to a new and a better technology named Gi-Fi.

Gi-Fi or Gigabit wireless [8] is the world's first transceiver integrated on a single chip that operates at 60GHz on the CMOS process. Gi-Fi allows wireless transfer of audio and video data up to 5 gigabits per second, usually within a range of 10 meters. Gi-Fi chip is a low cost, low power and high broadband chip that can easily be embedded into devices. And will be helpful for networking in homes and offices without wire.

Keywords*: Gi-fi, Wi-fi, IrDA, 802.15.3c.*

1. INTRODUCTION

Wireless communication is the transfer of information between two or more points that are not physically connected or as far as thousands or even millions of kilometres.

Wireless operations permit services, such as long range communications, that are impossible or impractical to implement with the use of wires. The term is commonly used in the telecommunications industry to refer to telecommunications systems (e.g. radio transmitters and receivers, remote controls, computer networks, network terminals, etc.) which use some form of energy (e.g. radio frequency (RF),acoustic energy, etc.) to transfer information without the use of wires. Information is transferred in this manner over both short and long distances. Many developments have done in the field of wireless for last 2-3 decades. Wi-Fi (ieee-802.11b) and WiMax (ieee-802.16e) have captured our attention more than other technologies like Bluetooth and Infrared because they are very slow and short distance technologies. As there is no recent developments which transfer data at faster rate, as video information transfer taking lot of time.

This leads to introduction of Gi-Fi technology. It offers some advantages over Wi-Fi, a similar wireless technology, in that it offers faster information rate in Gbps, less power consumption and low cost for short range transmissions.

Gi-Fi which is developed on an integrated wireless transceiver chip. In which a small antenna used and both transmitter- receiver integrated on a single chip, which is fabricated using the complementary metal oxide semiconductor (CMOS) process. Because of Gi-Fi transfer of large videos, files will be within seconds.

Other wireless technologies used for communication

1. Infrared

Infrared or **IrDA** is one of the earliest wireless technology used to provide wireless connectivity for devices that would normally use cables to connect. IrDA is a point-to-point, narrow angle (30° cone), ad-hoc data transmission standard designed to operate over a distance of 0 to 1 meter and at speeds of 9600 bps. IrDA is not able to penetrate solid objects and has limited data exchange applications compared to other wireless technologies.

2. Bluetooth

Bluetooth wireless technology is geared towards voice and data applications. Bluetooth wireless technology operates in the unlicensed 2.4 GHz spectrum. Bluetooth wireless technology can operate over a distance of 10 meters depending on the Bluetooth device class.

Bluetooth wireless technology is able to penetrate solid objects; it is the main advantage of this technology. Bluetooth technology is Omni-directional and does not require line-of-sight positioning of connected devices. Security has always been and continues to be a priority in the development of the Bluetooth specification. The Bluetooth also provide security.

3. Wi-Fi (ieee-802.11b)

"Wireless fidelity", Wi-Fi is one of the most popular wireless communication standards on the market. Wi-Fi technology was almost solely used to wirelessly connect laptop computers to the Internet via local area networks (LANs).Wi-Fi technology is now found in a host of non-computer electronic devices as well, such as home theatre receivers, video game consoles, Blue-ray players, digital cameras, and even GPS devices.

The Wi-Fi Alliance tests and certifies 802.11 based wireless equipment.

Some of the Wi-Fi standards are as follows:-802.11a: This uses OFDM, operates in the 5 GHz range, and has a maximum data rate of 54 Mbps.

802.11b: Operates in the 2.4 GHz range, has a maximum data rate of 11 Mbps and uses DSSS. 802.11b is the original Wi-Fi standard.

802.11g: Operates in the 2.4 GHz range, uses OFDM and has a maximum data rate of 54 Mbps. This is backwards compatible with 802.11b.

802.11e: This standard will improve quality of service**.**

4. WiMax

WiMax is a wireless metropolitan area network (MAN) technology. WiMax has a range of 50 km with data rates of 70 Mbps. Typical cell has a shorter range. The original 802.16 standard operated in the 10-66 GHz frequency bands with line of sight environments. The newly completed **802.16a** standard operates between 2 and 11 GHz and does not need line of sight. Delays in regulatory approval in Europe due to issues regarding the use of the spectrums in the 2.8 GHz and 3.4 GHz range. Supports vehicle mobility between 20 to 100+ km/hr.

5. Gi-Fi –

Gi-Fi or gigabit wireless is the world's first transceiver integrated on a single chip that operates at 60GHz on the CMOSprocess. It will allow wireless transfer of audio and video data at up to 5 gigabits per second [6], ten times the current maximum wireless transfer rate, at one-tenth the cost. NICTA researchers have chosen to develop this technology in the 57-64GHz unlicensed frequency band as the millimetre-wave range of the spectrum makes possible high component on-chip integration as well as allowing for the integration of very small high gain arrays. The available 7GHz of spectrum results in very high data rates, up to 5 gigabits per second to users within an indoor environment, usually within a range of 10 metres. It satisfies the standards of IEEE 802.15.3C.

Bluetooth Vs Wi-FI Vs Gi-Fi

Characteristic	**Bluetooth**	**Wi-Fi**	**Gi-Fi**
Frequency	2.4 GHz	2.4 GHz	7 GHz
Range	10 meters	100 meters	10 meters
Primary application	WPAN: Cable replacement	WLAN: Ethernet	Wireless PAN
Data transfer rate	800 Kbps	11 Mbps	5 Gbps
Power consumption	Low	Medium	Very Low
Specification authority	Bluetooth SIG	IEEE, WECA	IEEE

Table1: Comparison between Bluetooth, Wi-Fi & Gi-Fi

Architecture of Gi-Fi:

The core components of a GI-FI system is the subscriber station which available to several access points. It supports standard of IEEE 802.15.3C supports millimetre-wave wireless pan networks. The wireless PAN [1] is computer network used for communication among computer devices (including telephones and personal digital assistants) close to one person. An 802.15.3c based system often uses small antenna at the subscriber station. The antenna is mounted on the roof. It supports line of sight operation.

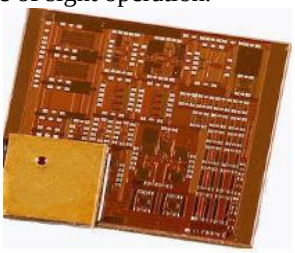


Fig1*:* The Gi-Fi integrated wireless transceiver chip developed at the National ICT Research Centre, Australia.

Fundamental Technologies in 802.15.3C [4]:

This mm Wave WPAN will operate in the new and clear band including 57-64 GHz unlicensed band defined by FCC 47 CFR 15.255. The millimetre-wave WPAN will allow high coexistence (close physical spacing) with all other microwave systems in the 802.15[5] family of WPANs.

60 GHz Band –

In this technology millimetre wave antenna[9] is used which operates at 60Ghz[2] frequency which is unlined band .Because of this band we are achieving high data rates energy propagation in the 60 GHz band has unique characteristics that make possible many other benefits such as excellent immunity to co-channel interference, high security, and frequency re-use. Point-to-point wireless systems operating at 60 GHz have been used for many years for satellite-to-satellite communications. This is because of high oxygen absorption at 60 GHz (10-15 dB/Km) [3]. This absorption attenuates.

60GHz [2] signals over distance, so that signals cannot travel far beyond their intended recipient. For this

reason, 60GHz [7] is an excellent choice for covert communications.

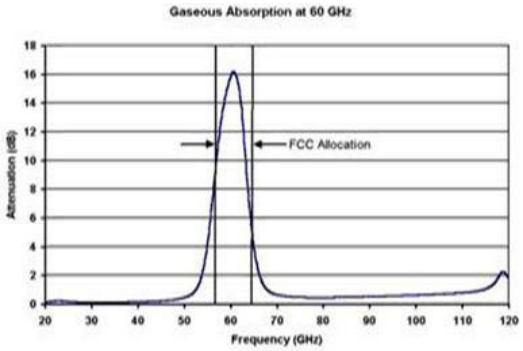


Fig2: Oxygen Attenuation vs. Frequency

Ultra Wide Band Frequency Usage:

UWB, a technology with high bit rate, high security and faster data transmission. It is a zero carrier technique with low coverage area. So we have low power consumption. These features are Ultra-Wideband (UWB) is a technology for transmitting information spread over a large bandwidth (>500 MHz) that should, be able to share spectrum with other users. Regulatory settings of FCC are intended to provide an efficient use of scarce radio bandwidth while enabling both high data rate personal-area network (PAN)[1] wireless connectivity and longer-range, low data rate applications as well as radar and imaging systems.

Features of Gi-Fi:

The Gi-Fi standard has been developed with many objectives in mind. These are summarized below:

1. High speed of data transfer

The main invention of Gi-Fi to provide higher bit rate .As the name itself indicates data transfer rate is in Giga bits per second. Speed of Gi-Fi is **5 gbps**, which is 10 times the present data transfer. Because of wider availability of continuous 7 GHz spectrum results in high data rates.

2. Low Power Consumption

As the large amount of information transfer it utilises mili watts of power only. It consumes only 2 mwatt power for data transfer of gigabits of information, where as in present technologies it takes 10 mwatt powers, which is very high.

3. High Security

Point-to-point wireless systems operating at 60 GHz have been used for many years by the intelligence community for high security communications and by the military for satellite-to satellite communications.

The combined effects of O2 absorption and narrow beam spread result in high security and low interference.

4. Cost-effective

Gi-Fi is based on an open, international standard. Mass adoption of the standard, and the use of low-cost, mass-produced chipsets, will drive costs down dramatically, and the resultant integrated wireless transreceiver chip which transfers data at high speed low power at low price $10 only. which is very less As compare to present systems .As go on development the price will be decreased.

Other features:

High level of frequency re-use enabled- Communication needs of multiple customers within a small geographic region can be satisfied.

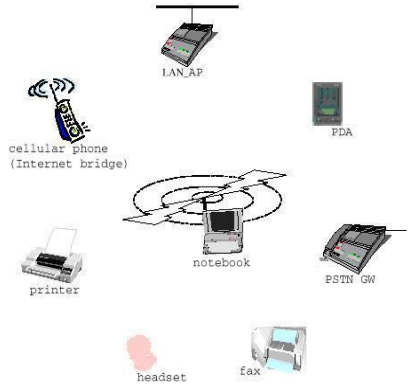
It is also highly portable-we can construct where ever we want.

It deploys line of sight operation having only shorter coverage area, it has more flexible architecture.

Applications

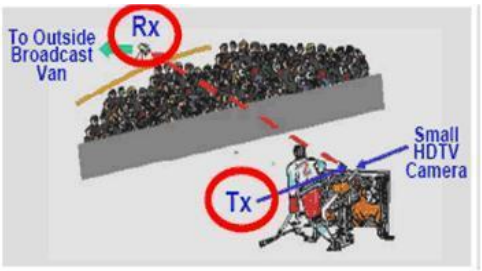
There are many usage scenarios that can be addressed by Gi-Fi. The following are some mobility usage applications of Gi-Fi.

In wireless PAN networks [6]

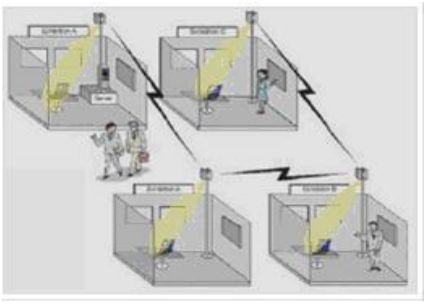


Inter-vehicle communication system

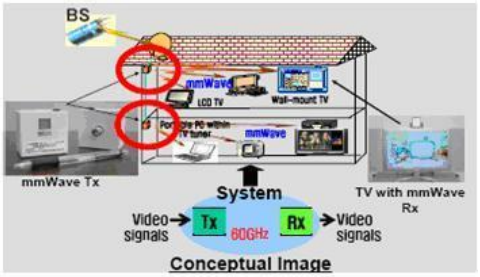
Broadcasting video signal transmission system in sports stadium



Ad-hoc information distribution with Point-to-Point network extension



mm-Wave video-signals transmission system



Huge data file transmission

It will transfer gigabits of information within seconds.

In Future

As the range is limited to shorter distances only we can expect the broad band with same speed and low power consumption.

Conclusion

Within five years, we expect Gi-Fi to be the dominant technology for wireless networking. By that time it will be fully mobile, as well as providing low-cost, high broadband access, with very high speed large files swapped within seconds which will develop wireless

home and office of future.

If the success of Wi-Fi and the imminent wide usage of WiMAX is any indication, Gi-Fi potentially can bring wireless broadband to the enterprise in an entirely new way.

REFERENCES

[1] IEEE 802.15 TG3c Working Group: Part 15.3: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for High Data Rate Wireless Personal Area Networks, (WPANs), 2009.

[2] IEEE 802.11 TGad Working Group: Draft Standard for Information Technology, Telecommunications and Information Exchange Between Systems, Local and Metropolitan Area Networks, Specific Requirements, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, Amendment 6: Enhancements for Very High Throughput in the60GHzBand, 2009.

[3] Jacob, M. and K¨urner, T.: IEEE 802.11 TGad Working Group: Measurement campaign at 60 GHz in the living room environment at TUBS, , 2009a.

[4] Funada, R., Harada, H., Shoji, Y., Kimura, R., Nishiguchi, Y., Lei,M., Choi, C.-S., Kojima F., Pyo, C. W., Lan, Z., Lakkis, I., Umehira, M., and Kato, S.: A design of single carrier based PHY for IEEE 802.15.3c standard, IEEE PIMRC, 1–5, 2007.

[5] Lei, M., Lakkis, I., Harada, H., and Kato, S.: MMSE-FDE based on Estimated SNR for Single-Carrier Block Transmission (SCBT)in Multi-Gbps WPAN (IEEE 802.15.3c), IEEE International Conference on Communications Workshop, 52–56, 2008.

[6] M. Lei, I. Lakkis, H. Harada, and S. Kato, "MMSE-FDE based on estimated SNR for single-carrier block transmission (SCBT) in multi- Gbps WPAN," in *Proc. IEEE ICC 2008*, pp.52–56, May 2008.

[7] M. Lei, C.-S. Choi, R. Funada, H. Harada, and S. Kato, "Throughput comparison of multi-Gbps WPAN PHY layer designs under non-linear 60-GHz power amplifier," in *Proc. IEEE PIMRC 2007*, pp.1–5, Sept. 2007

[8] K¨urner, T. and Jacob, M.: Application of ray tracing to derive channel models for future multi-gigabit systems, International Conference onElectromagnetics in Advanced Applications, 517– 520, 2009.

[9] Jacob, M. and K¨urner, T.: Radio channel characteristics for broadband WLAN/WPAN applications between 67 and 110 GHz, 3rd European Conference on Antennas and Propagation, 2663–2667, 2009b.

[10] ECMA-378 Standard, High Rate 60 GHz PHY, MAC and HDMI PAL, , 2008.

Internet of Things
Future Internet

Abstract

"Today computers—and, therefore, the Internet—are almost wholly dependent on human beings for information.. Conventional diagrams of the Internet … leave out the most numerous and important routers of all - people. The problem is, people have limited time, attention and accuracy—all of which means they are not very good at capturing data about things in the real world. And that's a big deal. We're physical, and so is our environment … You can't eat bits, burn them to stay warm or put them in your gas tank. Ideas and information are important, but things matter much more. Yet today's information technology is so dependent on data originated by people that our computers know more about ideas than things. If we had computers that knew everything there was to know about things—using data they gathered without any help from us—we would be able to track and count everything, and greatly reduce waste, loss and cost. We would know when things needed replacing, repairing or recalling, and whether they were fresh or past their best. The Internet of Things has the potential to change the world, just as the Internet did. Maybe even more so. "

The Internet of Things (IoT) is defined in many different ways, and it encompasses many aspects of life from connected homes and cities to connected cars and roads, roads to devices that track an individual s behavior and use the data collected for push services. Some mention one trillion Internet-connected devices by 2025 and define mobile phones as the eyes and ears of the applications connecting all of those connected things. By these internet of things billions objects can communicate over worldwide over a public, private internet protocol network in 2010, the number of everyday physical objects and devices connected to the Internet was around 12.5 billion. Smart cities, Smart cars, Public safety, Smart Industries and Environmental Protection has been given the high intention for future protection by IoT Ecosystem. For the development the government of Europe, Asia and America has considered the Internet of Things has area innovation and growth. Many visionaries have seized on the phrase Internet of Things to refer to the general idea of things, especially everyday objects, that are readable, recognizable, locatable, addressable, and/or controllable via the Internet, irrespective of the communication means (whether via RFID, wireless LAN, wide- area networks, or other means).. Due to internet of things hospitals are shifting to remote self-monitoring for patients. Internet of Things (IoT) is a new revolution of the Internet. Internet of Things (IoT) is can be said the expansion of internet services. It provides a platform for communication between objects where objects can organize and manage themselves. It makes objects themselves recognizable. The internet of things allows everyone to be connected anytime and anywhere. Objects can be communicated between each other by using radio frequency identification (RFID), wireless sensor network (WSN) ,etc. Radio Frequency identification assigns a unique identification to the objects. RFID technology is used as more secure identification and for tracking/locating objects, things, vehicle.

1.0 INTRODUCTION

Internet of Things (IoT) is a new revolution of the Internet. Internet of Things (IoT) is can be said the expansion of internet services. It provides a platform for communication between objects where objects can organize and manage themselves. It makes objects themselves recognizable. The internet of things allows

everyone to be connected any time and anywhere . RFID technology is used as more secure identification and for tracking/locating objects, things, vehicles.

Anyone who says that the Internet has fundamentally changed society may be right, but at the same time, the greatest transformation actually still lies ahead of us. Several new technologies are now converging in a way that means the Internet is on the brink of a substantial expansion as objects large and small get connected and assume their own web identity.

The Internet of Things (IoT) is the network of physical objects, devices, vehicles, buildings and other items which are embedded with electronics, software, sensors, and network connectivity, which enables these objects to collect and exchange data.The Internet of Things allows objects to be sensed and controlled remotely across existing network infrastructure, creating opportunities for more-direct integration between the physical world and computer-based systems, and resulting in improved efficiency, accuracy and economic benefit; when IoT is augmented with sensors and actuators, the technology becomes an instance of the more general class of cyber-physical systems, which also encompasses technologies such as smart grids, smart homes, intelligent transportation and smart cities. Experts estimate that the IoT will consist of almost 50 billion objects by 2020.

The interconnection of these embedded devices (including smart objects), is expected to usher in automation in nearly all fields, while also enabling advanced applications like a Smart Grid, and expanding to the areas such as smart cities.

"Things," in the IoT sense, can refer to a wide variety of devices such as heart monitoring implants, biochip transponders on farm animals, electric clams in coastal waters, automobiles with built-in sensors and rescue operations. These devices collect useful data with the help of various existing technologies and then autonomously flow the data between other devices. Current market examples include smart thermostat systems and washer/dryers that use Wi-Fi for remote monitoring.

IoT is also expected to generate large amounts of data from diverse locations that is aggregated very quickly, thereby increasing the need to better index, store and process such data. IoT is one of the platforms of today's Smart City and Smart Energy Management Systems.

2.0 Internet of Things:

'Internet of Things' refers to an era where things can be connected to each other using internet.

3.0 IMPACT OF INTERNET

The uses of internet include but not limited to usage of search engines which will help you to collect data from all over the world, usage of email and other instant message services which are giving flexibility of sharing information among groups within seconds, usage of internet in shopping via online shopping carts helped both clients and customers. Internet has become a platform to share knowledge between different communities. Several universities are publishing their research papers in their websites/digital libraries and helping other university students, researchers and professors scholar activities.

4.0 INTERNET USAGE

IDC estimates Internet of Things (IoT) market to grow to $8.9 trillion with over 212 billion connected things by 2020. The no. of connected devices surpassed total world population in year 2005 and it is estimated that no. of devices will be around 50 billion which is about 7 times of the world population at that time.

From the simplest day to day activities to the most complex human emotions, IoT will impact it.

5.0 Conclusion

The thought of always being tracked and your data being recorded does bring a fear to a consumer's mind, but we have to move away from it to see the benefits that this great technology is going to bring to us. The above examples were about a 'connected you', making your life seamless, but it brings with it higher benefits like connected cities, better commerce and an improved ecosystem.

As often happens, history is repeating itself. Just as in the early days when Cisco's tagline was "The Science of Networking Networks," IoT is at a stage where disparate networks and a multitude of sensors must come together and interoperate under a common set of standards.

This effort will require businesses, governments, standards organizations, and academia to work together toward a common goal. Next, for IoT to gain acceptance among the general populace, service providers and others must deliver applications that bring tangible value to peoples' lives. IoT must not represent the advancement of technology for technology's sake; the industry needs to demonstrate value in human terms.

In conclusion, IoT represents the next evolution of the Internet. Given that humans advance and evolve by turning data into information, knowledge, and wisdom, IoT has the potential to change the world as we know it today—for the better. How quickly we get there is up to us.

Suggested Readings

[1] https://en.wikipedia.org/wiki/Internet_of_Things

[2] www.wired.com/insights/2014/11/the-internet-of-things-bigger/

[3] http://internetofthingsagenda.techtarget.com/definition/Internet-of-Things-IoT

Phishing

Name : Azam khan
Institute of innovation in technology & management
New Delhi, India
anonymous.akey@gmail.com

Name : Harish Kumar Shukla
Institute of innovation in technology & management
New Delhi, India
Harishshukla817@gmail.com

Name : Abhishek Singh Shekhawat
Institute of innovation in technology & management
New Delhi, India
a.shekhawat1996@gmail.com

Abstract— Phishing is a kind of attack in which criminals use spoofed emails and fraudulent web sites to trick people into giving up personal information. This thesis looks at the phishing problem holistically by examining various stakeholders and their countermeasures, and by surveying experts' opinions about the current and future threats and the kinds of countermeasures that should be put in place. It composed of four studies. In the first study, we conducted semi-structured interviews with 31 anti-phishing experts from academia, law enforcement, and industry. We surveyed experts' opinions about the current and future of phishing threats and the kind of countermeasures that should be put in place. Our analysis led to eight key findings and 18 recommendations to improve phishing countermeasures. In the second study, we study the effectiveness of popular phishing tools that are used by major web browsers. We used fresh phish that were less than 30 minutes old to conduct two tests on eight anti-phishing toolbars. We found blacklists were ineffective when protecting users initially. The tools that uses heuristics to complement blacklists caught significantly more phish than black list only tools with very low false positives. In the third study, we describe the design and evaluation of Anti-Phishing Phil, an online game that teaches users good habits to help them avoid phishing attacks. We used learning science principles to design and iteratively refine the game. We evaluated Anti-Phishing Phil through laboratory and real-world experiments. These experiments showed that people trained with Anti-Phishing Phil were much better at detecting phishing websites, and they retain knowledge after one week. In the fourth and final study we present our results of a role play survey instrument administered to 1001 online survey respondents to study both the relationship between demographics and phishing susceptibility, and the effectiveness of several anti-phishing educational materials. Our results suggest that women are more susceptible than men to phishing iv and participants between the ages of 18 and 25 are more susceptible to phishing than other age groups. We explain these demographic factors through a mediation analysis. Educational materials reduced users tendency to enter information into phishing web pages by 40% percent; however, some of the educational materials we tested also slightly decreased participants tendency to click on legitimate links.

Keywords
Phishing, Social engineering, Fraud, Spam, Security

Introduction

Phishing is the attempt to obtain sensitive information such as usernames, passwords, and credit card details (and, indirectly, money), often for malicious reasons, by disguising as a trustworthy entity in an electronic communication.[1][2] The word is a neologism created as a homophone of fishing due to the similarity of using a bait in an attempt to catch a victim. According to the 3rd Microsoft Computing Safer Index Report released in February 2014, the annual worldwide impact of phishing could be as high as $5 billion.[3]

Phishing is typically carried out by email spoofing[4] or instant messaging,[5] and it often directs users to enter personal information at a fake website, the look and feel of which are almost identical to the legitimate one. Communications purporting to be from social web sites, auction sites, banks, online payment processors or IT administrators are often used to lure victims. Phishing emails may contain links to websites that are infected with malware.[6]

EASE OF USE(TYPE OF PHISHING)

Phishing is an example of social engineering techniques used to deceive users, and exploits weaknesses in current web security.[7] Attempts to deal with the growing number of reported phishing incidents include legislation, user training, public awareness, and technical security measures. Many websites have now created secondary tools for applications, like maps for games, but they should be clearly marked as to who wrote them, and users should not use the same passwords anywhere on the internet.

Techniques

Phishing types

Spear phishing

Phishing attempts directed at specific individuals or companies have been termed spear phishing.[8] Attackers may gather personal information about their target to increase their probability of success. This technique is, by far, the most successful on the internet today, accounting for 91% of attacks.[9]

Clone phishing

Clone phishing is a type of phishing attack whereby a legitimate, and previously delivered, email containing an attachment or link has had its content and recipient address(es) taken and used to create an almost identical or cloned email. The attachment or link within the email is replaced with a malicious version and then sent from an email address spoofed to appear to come from the original sender. It may claim to be a resend of the original or an updated version to the original. This technique could be used to pivot (indirectly) from a previously infected machine and gain a foothold on another machine, by exploiting the social trust associated with the inferred connection due to both parties receiving the original email.

Whaling

Several phishing attacks have been directed specifically at senior executives and other high-profile targets within businesses, and the term whaling has been coined for these kinds of attacks.[10] In the case of whaling, the masquerading web page/email will take a more serious executive-level form. The content will be crafted to target an upper manager and the person's role in the company. The content of a whaling attack email is often written as a legal subpoena,

customer complaint, or executive issue. Whaling scam emails are designed to masquerade as a critical business email, sent from a legitimate business authority. The content is meant to be tailored for upper management, and usually involves some kind of falsified company-wide concern. Whaling phishermen have also forged official-looking FBI subpoena emails, and claimed that the manager needs to click a link and install special software to view the subpoena.[11]

Link manipulation

Most methods of phishing use some form of technical deception designed to make a link in an email (and the spoofed website it leads to) appear to belong to the spoofed organization.[12]Misspelled URLs or the use of subdomains are the common tricks used by phishers. In the following example URL, http://www.yourbank.example.com/, it appears as though the URL will take you to the example section of the yourbank website; actually this URL points to the "yourbank" (i.e. phishing) section of the example website. Another common trick is to make the displayed text for a link (the text between the <A> tags) suggest a reliable destination, when the link actually goes to the phishers' site. Many email clients or web browsers will show previews of where a link will take the user in the bottom left of the screen, while hovering the mouse cursor over a link.[13] This behaviour, however, may in some circumstances be overridden by the phisher.

A further problem with URLs has been found in the handling of internationalized domain names (IDN) in web browsers, that might allow visually identical web addresses to lead to different, possibly malicious, websites. Despite the publicity surrounding the flaw, known as IDN spoofing[14] or homograph attack,[15] phishers have taken advantage of a similar risk, using open URL redirectors on the websites of trusted organizations to disguise malicious URLs with a trusted domain.[16][17][18] Even digital certificates do not solve this problem because it is quite possible for a phisher to purchase a valid certificate and subsequently change content to spoof a genuine website, or, to host the phish site without SSL at all.[19]

Filter evasion

Phishers have even started using images instead of text to make it harder for anti-phishing filters to detect text commonly used in phishing emails.[20] However, this has led to the evolution of more sophisticated anti-phishing filters that are able to recover hidden text in images. These filters use OCR (optical character recognition) to optically scan the image and filter it.[21] Some anti-phishing filters have even used IWR (intelligent word recognition), which is not meant to completely replace OCR, but these filters can even detect cursive, hand-written, rotated (including upside-down text), or distorted (such as made wavy, stretched vertically or laterally, or in different directions) text, as well as text on colored backgrounds.

Website forgery

Once a victim visits the phishing website, the deception is not over. Some phishing scams use JavaScript commands in order to alter the address bar.[22] This is done either by placing a picture of a legitimate URL over the address bar, or by closing the original bar and opening up a new one with the legitimate URL.[23]

An attacker can even use flaws in a trusted website's own scripts against the victim.[24] These types of attacks (known as cross-site scripting) are particularly problematic, because they direct the user to sign in at their bank or service's own web page, where everything from the web address to the security certificates appears correct. In reality, the link to the website is crafted to carry out the attack, making it very difficult to spot without specialist knowledge. Just such a flaw was used in 2006 against PayPal.[25]

A Universal Man-in-the-middle (MITM) Phishing Kit, discovered in 2007, provides a simple-to-use interface that allows a phisher to convincingly reproduce websites and capture log-in details entered at the fake site.[26]

To avoid anti-phishing techniques that scan websites for phishing-related text, phishers have begun to use Flash-based websites (a technique known as phlashing). These look much like the real website, but hide the text in a multimedia object.[27]

Covert redirect

Covert redirect is a subtle method to perform phishing attacks that makes links appear legitimate, but actually redirect a victim to an attacker's website. The flaw is usually masqueraded under a log-in popup based on an affected site's domain.[28] It can affect OAuth 2.0 and OpenID based on well-known exploit parameters as well. This often makes use of open redirect and XSS vulnerabilities in the third-party application websites.[29]

Normal phishing attempts can be easy to spot because the malicious page's URL will usually be different from the real site link. For covert redirect, an attacker could use a real website instead by corrupting the site with a malicious login popup dialogue box. This makes covert redirect different from others.[30][31]

For example, suppose a victim clicks a malicious phishing link beginning with Facebook. A popup window from Facebook will ask whether the victim would like to authorize the app. If the victim chooses to authorize the app, a "token" will be sent to the attacker and the victim's personal sensitive information could be exposed. These information may include the email address, birth date, contacts, and work history.[29] In case the "token" has greater privilege, the attacker could obtain more sensitive information including the mailbox, online presence, and friends list. Worse still, the attacker may possibly control and operate the user's account.[32] Even if the victim does not choose to authorize the app, he or she will still get redirected to a website controlled by the attacker. This could potentially further compromise the victim.[33]

This vulnerability was discovered by Wang Jing, a Mathematics Ph.D. student at School of Physical and Mathematical Sciences in Nanyang Technological University in Singapore.[34]

Covert redirect is a notable security flaw, though it is not a threat to the Internet worth significant attention.[35]

Social engineering

Users can be incentivised to click on various kinds of unexpected content for a variety of technical and social reasons. For example, a malicious attachment might masquerade as a benign linked Google doc.[36]

Alternatively users might be outraged by a fake news story, click a link and become infected.[37]

Phone phishing

Not all phishing attacks require a fake website. Messages that claimed to be from a bank told users to dial a phone number regarding problems with their bank accounts.[38] Once the phone number (owned by the phisher, and provided by a voice over IP service) was dialed, prompts told users to enter their account numbers and PIN. Vishing (voice phishing) sometimes uses fake caller-ID data to give the appearance that calls come from a trusted organisation.[39] SMS phishing uses cell phone text messages to induce people to divulge their personal information.

History

1980s

A phishing technique was described in detail in a paper and presentation delivered to the 1987 International HP Users Group, Interex.[42]

1990s

The term 'phishing' is said to have been coined by the well known spammer and hacker in the mid-90s, Khan C Smith.[43] The first recorded mention of the term is found in the hacking tool AOHell (according to its creator), which included a function for attempting to steal the passwords or financial details of America Online users.[44][45]

Early AOL phishing

Phishing on AOL was closely associated with the warez community that exchanged unlicensed software and the black hat hacking scene that perpetrated credit card fraud and other online crimes. AOL enforcement would detect words used in AOL chat rooms to suspend the accounts individuals involved in counterfeiting software and trading stolen accounts. The term was used because '<><' is the single most common tag of HTML that was found in all chat transcripts naturally, and as such could not be detected or filtered by AOL staff. The symbol <>< was replaced for any wording that referred to stolen credit cards, accounts, or illegal activity. Since the symbol looked like a fish, and due to the popularity of phreaking it was adapted as 'Phishing'. AOHell, released in early 1995, was a program designed to hack AOL users by allowing the attacker to pose as an AOL staff member, and send an instant message to a potential victim, asking him to reveal his password.[46] In order to lure the victim into giving up sensitive information, the message might include imperatives such as "verify your account" or "confirm billing information". Once the victim had revealed the password, the attacker could access and use the victim's account for fraudulent purposes. Both phishing and warezing on AOL generally required custom-written programs, such as AOHell. Phishing became so prevalent on AOL that they added a line on all instant messages stating: "no one working at AOL will ask for your password or billing information", though even this didn't[tone] prevent some people from giving away their passwords and personal information if they read and believed the IM first. A user using both an AIM account and an AOL account from an ISP simultaneously could phish AOL members with relative impunity as internet AIM accounts could be used by non-AOL internet members and could not be auctioned (i.e., reported to AOL TOS department for disciplinary action).[47][tone]. In late 1995, AOL crackers resorted to phishing for legitimate accounts after AOL brought in measures in late 1995 to prevent using fake, algorithmically generated credit card numbers to open accounts.[48] Eventually, AOL's policy enforcement forced copyright infringement off AOL servers, and AOL promptly deactivate accounts involved in phishing, often before the victims could respond. The shutting down of the wares scene on AOL caused most phishes to leave the service.[49]

2000s

2001

The first known direct attempt against a payment system affected E-gold in June 2001, which was followed up by a "post-9/11 id check" shortly after the September 11 attacks on the World Trade Center.[50]

2003

The first known phishing attack against a retail bank was reported by The Banker in September 2003.[51]

2004

It is estimated that between May 2004 and May 2005, approximately 1.2 million computer users in the United States suffered losses caused by phishing, totaling approximately US$929 million. United States businesses lose an estimated US$2 billion per year as their clients become victims.[52]

Phishing is recognized as a fully organized part of the black market. Specializations emerged on a global scale that provided phishing software for payment (thereby outsourcing risk), which were assembled and implemented into phishing campaigns by organized gangs.[53][54]

2005

In the United Kingdom losses from web banking fraud—mostly from phishing—almost doubled to GB£23.2m in 2005, from GB£12.2m in 2004,[55] while 1 in 20 computer users claimed to have lost out to phishing in 2005.[56]

2006

Almost half of phishing thefts in 2006 were committed by groups operating through the Russian Business Network based in St. Petersburg.[57]

Banks dispute with customers over phishing losses. The stance adopted by the UK banking body APACS is that "customers must also take sensible precautions ... so that they are not vulnerable to the criminal."[58] Similarly, when the first spate of phishing attacks hit the Irish Republic's banking sector in September 2006, the Bank of Ireland initially refused to cover losses suffered by its customers,[59] although losses to the tune of €113,000 were made good.[60]

Phishes are targeting the customers of banks and online payment services. Emails, supposedly from the Internal Revenue Service, have been used to glean sensitive data from U.S. taxpayers.[61] While the first such examples were sent indiscriminately in the expectation that some would be received by customers of a given bank or service, recent research has shown that phishers may in principle be able to determine which banks potential victims use, and target bogus emails accordingly.[62]

Social networking sites are a prime target of phishing, since the personal details in such sites can be used in identity theft;[63] in late 2006 a computer worm took over pages on MySpace and altered links to direct surfers to websites designed to steal login details.[64] Experiments show a success rate of over 70% for phishing attacks on social networks.[65]

2007

3.6 million adults lost US$3.2 billion in the 12 months ending in August 2007.[66] Microsoft claims these estimates are grossly exaggerated and puts the annual phishing loss in the US at US$60 million.[67]

Attackers who broke into TD Ameritrade's database and took 6.3 million email addresses (though they were not able to obtain social security numbers, account numbers, names, addresses, dates of birth, phone numbers and trading activity) also wanted the account usernames and passwords, so they launched a follow-up spear phishing attack.[68]

2008

The Rapid Share file sharing site has been targeted by phishing to obtain a premium account, which removes speed caps on downloads, auto-removal of uploads, waits on downloads, and cool down times between uploads.[69]

Crypto currencies such as Bit coin, introduced in late 2008, facilitate the sale of malicious software, making transactions secure and anonymous.

2009

In January 2009, a phishing attack resulted in unauthorized wire transfers of US$1.9 million through Expert-Metal's online banking accounts.

In the 3rd Quarter of 2009, the Anti-Phishing Working Group reported receiving 115,370 phishing email reports from consumers with US and China hosting more than 25% of the phishing pages each.

REFERENCES

Anti-Phishing Working Group (2009) "Phishing Activity Trends Report: 3rd Quarter 2009"
Available at: http://www.antiphishing.org/reports/apwg_report_Q3_2009.pdf(Accessed: 15 January2010).

Dhamija, R., Tygar, J.D. and Hearst, M. (2006) "Why Phishing Works", Proceedings of the SIGCHI conference on Human Factors in computing systems,
Montréal, Québec, Canada. pp. 581-590.

Gartner (2010) "Gartner Survey Shows Phishing Attacks Escalated in 2007; More than $3 Billion Lost to These Attacks", Available at: http://www.gartner.com/it/page.jsp?id=565125 (Accessed: 21 January 2010).

Herzberg, A. and Jbara, A. (2008) "Security and Identification Indicators for Browsers against Spoofing and Phishng Attacks" ACM Transaction on Internet Technology, 8 (4)16, pp.1-36.

Jagatic, T.N., Johnson, N.A., Jakobsson, M., Menezer, F. (2007) "Social Phishing", Communications of the ACM, 50 (10), pp. 94-100.

Karakasiliotis, A., Furnell, S.M. and Papadaki, M. (2007) "An assessment of end-user vulnerability to phishing attacks", Journal of Information Warfare, 6 (1), pp. 17-28.Krebs, B. (2004)
"Phishing Schemes Scar Victims", Available at: http://www.washingtonpost.com/ac2/wp-dyn/A59349
2004Nov18?language=printer (Accessed: 16 June 2010).

Research on Future Internet Architectures and It's Protocol

Chirag Chopra

Institute of innovation in technology & management
New Delhi, India
Chirag0814@gmail.com

Sandeep Rawat

Institute of innovation in technology & management
New Delhi, India
Sandeepraawat169@gmail.com

Rohan Chhabra

Institute of innovation in technology & management
New Delhi, India
Chhabra.rohan4@gmail.com

ABSTRACT
The current Internet, which was designed over 40 years ago, is facing unprecedented challenges in many aspects, especially in the commercial context. The emerging demands for security, mobility, content distribution, etc. are hard to be met by incremental changes through ad-hoc patches. New clean-slate architecture designs based on new design principles are expected to address these challenges. In this survey article, we investigate the key research topics in the area of future Internet architecture. Many ongoing research projects from United States, the European Union, Japan, China, and other places are introduced and discussed. We aim to draw an overall picture of the current research progress on the future Internet architecture.

I. INTRODUCTION

The Internet has evolved from an academic network to a broad commercial platform. It has become an integral and indispensable part of our daily life, economic operation, and society. However, many technical and non-technical challenges have emerged during this process, which call for potential new Internet architectures. Technically, the current Internet was designed over 40 years ago with certain design principles. Its continuing success has been hindered by more and more sophisticated network attacks due to the lack of security embedded in the original architecture. Also, IP's narrow waist means that the core architecture is hard to modify, and new functions have to be implemented through myopic and clumsy ad hoc patches on top of the existing architecture. Moreover, it has become extremely difficult to support the ever increasing demands for security, performance reliability, social content distribution, mobility, and so on through such incremental changes. As a result, a clean-slate architecture design paradigm has been suggested by the research community to build the future Internet. From a non-technical aspect, commercial usage requires fine-grained security enforcement as opposed to the current "perimeter-based" enforcement.

Security needs to be an inherent feature and integral part of the architecture. Also, there is a significant demand to transform the Internet from a simple "host-to-host" packet delivery paradigm into a more diverse paradigm built around the data, content, and users instead of the machines. All of the above challenges have led to the research on future Internet architectures.
Future Internet architecture is not a single improvement on a specific topic or goal. A cleanslate solution on a specific topic may assume the other parts of the architecture to be fixed and unchanged. Thus, assembling different cleanslate

solutions targeting different aspects will not necessarily lead to a new Internet architecture. Instead, it has to be an overall redesign of the whole architecture, taking all the issues (security, mobility, performance reliability, etc.) into consideration. It also needs to be evolvable and flexible to accommodate future changes. Most previous clean-slate projects were focused on individual topics. Through a collaborative and comprehensive approach, the lessons learned and research results obtained from these individual efforts can be used to build a holistic Internet architecture.

II. Key Research Topics

In this section, we discuss some key research topics that are being addressed by different research projects.

Content- or data-oriented paradigms: Today's Internet builds around the "narrow waist" of IP, which brings the elegance of diverse design above and below IP, but also makes it hard to change the IP layer to adapt for future requirements. Since the primary usage of today's Internet has changed from host-to-host communication to content distribution, it is desirable to change the architecture's narrow waist from IP to the data or content distribution. Several research projects are based on this idea. This category of new paradigms introduces challenges in data and content security and privacy, scalability of naming and aggregation, compatibility and co-working with IP, and efficiency of the new paradigm.

Mobility and ubiquitous access to networks: The Internet is experiencing a significant shift from PC-based computing to mobile computing. Mobility has become the key driver for the future Internet. Convergence demands are increasing among heterogeneous networks such as cellular, IP, and wireless ad hoc or sensor networks that have different technical standards and business models. Putting mobility as the norm instead of an exception of the architecture potentially nurtures future Internet architecture with innovative scenarios and applications. Many collaborative research projects in academia and industry are pursuing such research topics with great interest. These projects also face challenges such as how to trade off mobility with scalability, security, and privacy protection of mobile users, mobile endpoint resource usage optimization, and so on.

***Cloud-computing-centric architectures*:** Migrating storage and computation into the "cloud" and creating a "computing utility" is a trend that demands new Internet services and applications. It creates new ways to provide global-scale resource provisioning in a "utilitylike" manner. Data centers are the key components of such new architectures. It is important to create secure, trustworthy, extensible, and robust architecture to interconnect data, control, and management planes of data centers. The cloud computing perspective has attracted considerable research effort and industry projects toward these goals. A major technical challenge is how to guarantee the trustworthiness of users while maintaining persistent service availability.

Security: Security was added into the original Internet as an additional overlay instead of an inherent part of the Internet architecture. Now security has become an important design goal for the future Internet architecture. The research is related to both the technical context and the economic and public policy context. From the technical aspect, it has to provide multiple granularities (encryption, authentication, authorization, etc.) for any potential use case. Also, it needs to be open and extensible to future new security related solutions. From the non-technical aspect, it should ensure a trustworthy interface among the participants (e.g., users, infrastructure providers, and content providers). There are many research projects and working groups related to security. The challenges on this topic are very diverse, and multiple participants make the issue complicated.

***Experimental testbeds*:** As mentioned earlier, developing new Internet architectures requires large-scale testbeds. Currently, testbed research includes multiple testbeds with different virtualization technologies, and the federation and coordination among these testbeds. Research organizations from the United States, European Union, and Asia have initiated several programs related to the research and implementation of large-scale testbeds. These projects explore challenges related to large-scale hardware, software, distributed system test and maintenance, security and robustness, coordination, openness, and extensibility.

Besides these typical research topics, there are several others, including but not limited to networked multimedia; "smart dust," also called the "Internet of things"; and Internet services architecture. However, note that in this survey, we are not trying to enumerate all the possible topics and corresponding research projects. Instead, we focus on a representative subset and discuss a few important ongoing research projects.

Due to length limitations, we are not able to enumerate all the references for the projects discussed below. However, we do have a longer survey [18], which includes a more complete reference list for further reading.

Since the primary usage of the today's Internet has changed from host-to-host communication to content distribution, it is desirable to change the architecture's narrow waist from IP to the data or content distribution.

.

Named Data Networking (NDN)

The basic argument of the NDN[3] project is that the primary usage of the current Internet has changed from end-to-end packet delivery to a content-centric model. The current Internet, which is a "client-server" model, is facing challenges in supporting secure content-oriented functionality. In this information dissemination model, the network is "transparent" and just forwarding data (i.e., it is "content-unaware"). Due to this unawareness, multiple copies of the same data are sent between endpoints on the network again and again

without any traffic optimization on the network's part. The NDN uses a different model that enables the network to focus on "what" (contents) rather than "where" (addresses). The data are named instead of their location (IP addresses). Data become the first-class entities in NDN. Instead of trying to secure the transmission channel or data path through encryption, NDN tries to secure the content by naming the data through a security-enhanced method. This approach allows separating trust in data from trust between hosts and servers, which can potentially enable content caching on the network side to optimize traffic.

NDN has several key research issues. The first one is how to find the data, or how the data are named and organized to ensure fast data lookup and delivery. The proposed idea is to name the content by a hierarchical "name tree" which is scalable and easy to retrieve. The second research issue is data security and trustworthiness. NDN proposes to secure the data directly instead of securing the data "containers" such as files, hosts, and network connections. The contents are signed by public keys. The third issue is the scaling of NDN. NDN names are longer than IP addresses, but the hierarchical structure helps the efficiency of lookup and global accessibility of the data.

Regarding these issues, NDN tries to address them along the way to resolve the challenges in routing scalability, security and trust models, fast data forwarding and delivery, content protection and privacy, and an underlying theory supporting the design.

MobilityFirst —The[4] basic motivation of MobilityFirst is that the current Internet is designed for interconnecting fixed endpoints. It fails to address the trend of dramatically increasing demands of mobile devices and services. The Internet usage and demand change is also a key driver for providing mobility from the architectural level for the future Internet. For the near term, MobilityFirst aims to address the cellular convergence trend motivated by the huge mobile population of 4 to 5 billion cellular devices; it also provides mobile peer-to-peer (P2P) and infostation (delay-tolerant network [DTN]) application services which offer robustness in case of link/network disconnection. For the long term, in the future, MobilityFirst has the ambition of connecting millions of cars via vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) modes, which involve capabilities such as location services, georouting, and reliable multicast. Ultimately, it will introduce a pervasive system to interface human beings with the physical world, and build a future Internet around people.

The challenges addressed by MobilityFirst include stronger security and trust requirements due to open wireless access, dynamic association, privacy concerns, and greater chance of network failure. MobilityFirst targets a clean-slate design directly addressing mobility such that the fixed Internet will be a special case of the general design. MobilityFirst builds the "narrow waist" of the protocol stack around several protocols: • Global name resolution and routing service

- Storage-aware (DTN-like) routing protocol
- Hop-by-hop segmented transport
- Service and management application programming interfaces (APIs)

The DTN-like routing protocol is integrated with the use of self-certifying public key addresses for inherent trustworthiness. Functionalities such as context- and location-aware services fit into the architecture naturally. It shows all the building blocks mentioned above and how they work together.

eXpressive Internet Architecture (XIA) —As[5] we observe, most of the research projects on future Internet architectures realize the importance of security and consider their architecture carefully to avoid the flaws of the original Internet design. However, XIA directly and explicitly targets the security issue within its design. There are three key ideas in the XIA architecture:

- Define a rich set of building blocks or communication entities as network principals including hosts, services, contents, and future additional entities.
- It is embedded with intrinsic security by using self-certifying identifiers for all principals for integrity and accountability properties.
- A pervasive "narrow waist" (not limited to the host-based communication as in the current Internet) for all key functions, including access to principals, interaction among stakeholders, and trust management; it aims to provide interoperability at all levels in the system, not just packet forwarding.

The core of the XIA is the Expressive Internet Protocol (XIP) supporting communication between various types of principals. Three typical XIA principal types arecontent, host (defined by "who"), and service (defined by what it does). They are open to future extension. Each type of principal has a narrow waist that defines the minimal functionality required for interoperability. Principles talk to each other using expressive identifiers (XIDs), which are 160 bit identifiers identifying hosts, pieces of content, or services. The XIDs are basically self-certifying identifiers taking advantage of cryptographic hash technology. By using this XID, the content retrieval no longer relies on a particular host, service or network path. XIP can then support future functions as a diverse set of services. For low-level services, it uses a path-segment-based network architecture (named Tapa in their previous work) as the basic building block; and builds services for content-transfer and caching and service for secure content provenance at a higher level. XIA also needs various trustworthy mechanisms and provides network availability even when under attack. Finally, XIA defines explicit interfaces between network actors with different roles and goals.

GLOBAL ENVIRONMENT FOR NETWORK INNOVATIONS (GENI)

GENI [6] is a collaborative program supported by NSF aimed at providing a global large-scale experimental testbed for future Internet architecture test and validation. Started in 2005, it has attracted broad interest and participation from both academia and industry. Besides its initial support from existing projects on a

dedicated backbone network infrastructure, it also aims to attract other infrastructure platforms to participate in the federation — the device control framework to provide these participating networks with users and operating environments, to observe, measure, and record the resulting experimental outcomes. So generally, GENI is different from common testbeds in that it is a generalpurpose large-scale facility that puts no limits on the network architectures, services, and applications to be evaluated; it aims to allow clean-slate designs to experiment with real users under real conditions.

The key idea of GENI is to build multiple virtualized *slices* out of the substrate for resource sharing and experiments. It contains two key pieces:

- Physical network substrates that are expandable building block components
- A global control and management framework that assembles the building blocks together into a coherent facility

Thus, intuitively two kinds of activities will be involved in GENI testbeds: one is deploying a prototype testbed federating different small and medium ones together (e.g., the OpenFlow testbed for campus networks [8]); the other is to run observable, controllable, and recordable experiments on it.

There are several working groups concentrating on different areas, such as the control framework working group; GENI experiment workflow and service working group; campus/operation, management, integration, and security working group; and instrumentation and management working group.

The GENI generic control framework consists of several subsystems and corresponding basic entities: • Aggregate and components

- Clearinghouse
- Research organizations, including researchers and experiment tools
- Experiment support service
- "Opt-in" end users
- GENI operation and management

Clearinghouses from different organizations and places (e.g., those from the United States and European Union) can be connected through federation. By doing this, GENI not only federates with identical "GENI-like" systems, but also with any other system if they comply with a clearly defined and relatively narrow set of interfaces for federation. With these entities and subsystems, "slices" can be created on top of the shared substrate for miscellaneous researchdefined specific experiments, and end users can "opt in" onto the GENI testbed accordingly.

GENI's research and implementation plan consists of multiple continuous *spirals* (currently in spiral 3). Each spiral lasts for 12 months. Spiral 1 ran from 2008 to 2009; spiral 2 ran from 2009 to 2010; spiral 3 started in 2011. In spiral 1, the primary goals were to demonstrate one or more early prototypes of the GENI control framework and end-to-end slice operation across multiple technologies; there were five competing approaches to the GENI control framework, called "clusters."

Cluster A was the Trial Integration Environment based on DETER (TIED) control framework focusing on federation, trust, and security. It was a one-project cluster based on the CyberDefense Technology Experimental Research (DETER) control framework by the University of Southern California (USC)/ISI, which is an individual "mini-GENI" testbed to demonstrate federated and coordinated network provisioning. Cluster A particularly aimed to provide usability across

multiple communities through federation. The project delivered software "fedd" as the implementation of the TIED federation architecture providing dynamic and on-demand federation, and interoperability across ProtoGENI, GENIAPI, and non-GENI aggregate. It included an Attribute Based Access Control (ABAC) mechanism for large-scale distributed systems. It created a federation with two other projects: StarBED in Japan and ProtoGENI in the United States.

Cluster B was a control framework based on PlanetLab implemented by Princeton University emphasizing experiments with virtualized machines over the Internet. By the end of spiral 2, it included at least 12 projects from different universities and research institutes. The results of these projects are to be integrated into the PlanetLab testbed. PlanetLab provided "GENIwrapper" code for independent development of an aggregate manager (AM) for Internet entities. A special "lightweight" protocol was introduced to interface PlanetLab and OpenFlow equipment. Through these mechanisms, other projects in the cluster can design their own substrates and component managers with different capacities and features.

Cluster C was the ProtoGENI control framework by the University of Utah based on Emulab, emphasizing network control and management. By the end of spiral 2, it consisted of at least 20 projects. The cluster integrated these existing and under-construction systems to provide key GENI functions. The integration included four key components: a backbone based on Internet2; sliceable and programmable PCs and NetFPGA cards; and subnets of wireless and wired edge clusters. Cluster C so far is the largest set of integrated projects in GENI.

Cluster D was Open Resource Control Architecture (ORCA) from Duke University and RENCI focusing on resource allocation and integration of sensor networks. By the end of spiral 2, it consisted of five projects. ORCA tried to include optical resources from the existing Metro-Scale Optical Testbed (BEN). Different from other clusters, the ORCA implementation included the integration of wireless/sensor prototypes. It maintains a clearinghouse for the testbeds under the ORCA control framework through which it connects to the national backbone and is available to external researchers.

Cluster E was Open-Access Research Testbed for Next-Generation Wireless Networks (ORBIT) by Rutgers University focusing on mobile and wireless testbed networks. It included three projects by the end of spiral 2. The basic ORBIT did not include a full clearinghouse implementation. Cluster E tried to research how mobile and wireless work can affect and possibly be merged into the GENI architecture. WiMAX is one of the wireless network prototypes in this cluster.

In this article, we present a survey of the current research efforts on future Internet architectures. It is not meant to be a complete enumeration of all such projects. Instead, we focus on a series of representative research projects. Research programs and efforts from the United States, European Union, and Asia are discussed. By doing this, we hope to draw an approximate overall picture of the up-to-date status in this area.

III. SUMMARY

In this article, we present a survey of the current research efforts on future Internet architectures. It is not meant to be a complete enumeration of all such projects. Instead, we focus on a series of representative research projects. By doing this, we hope to draw an approximate overall picture of the up-to-date status in this area.

REFERENCES

[1] NSF Future Internet Architecture Project, http://www. nets-fia.net/.
[2] Named Data Networking Project, http://www.nameddata.net.
[3] MobilityFirst Future Internet Architecture Project, http://mobilityfirst.winlab.rutgers.edu/.

[4] eXpressive Internet Architecture Project, http://www.cs.cmu.edu/~xia/.
[5] Global Environment for Network Innovations (GENI) Project, http://www.geni.net/

A Survey on Future Internet Architectures and It's Protocol

Chirag Chopra
Chiraga0814@gmail.com
Institute of innovation in technology & management
New Delhi, India

Sandeep Rawat
Sandeeprawat169@gmail.com
Institute of innovation in technology & management
New Delhi, India

Rohan Chhabra
Chhabra.rohan4@gmail.com
Institute of innovation in technology & management
New Delhi, India

ABSTRACT
The current Internet, which was designed over 40 years ago, is facing unprecedented challenges in many aspects, especially in the commercial context. The emerging demands for security, mobility, content distribution, etc. are hard to be met by incremental changes through ad-hoc patches. New clean-slate architecture designs based on new design principles are expected to address these challenges. In this survey article, we investigate the key research topics in the area of future Internet architecture. Many ongoing research projects from United States, the European Union, Japan, China, and other places are introduced and discussed. We aim to draw an overall picture of the current research progress on the future Internet architecture.

Keywords:

Cloud-computing,Security,Privacy,Internet Protocol,Network,Mobility

IV. INTRODUCTION

The Internet has evolved from an academic network to a broad commercial platform. It has become an integral and indispensable part of our daily life, economic operation, and society. However, many technical and non-technical challenges have emerged during this process, which call for potential new Internet architectures. Technically, the current Internet was designed over 40 years ago with certain design principles. Its continuing success has been hindered by more and more sophisticated network attacks due to the lack of security embedded in the original architecture. Also, IP's narrow waist means that the core architecture is hard to modify, and new functions have to be implemented through myopic and clumsy ad hoc patches on top of the existing architecture. Moreover, it has become extremely difficult to support the ever increasing demands for security, performance reliability, social content distribution, mobility, and so on through such incremental changes. As a result, a clean-slate architecture design paradigm has been suggested by the research community to build the future Internet. From a non-technical aspect, commercial usage requires fine-grained security enforcement as opposed to the current "perimeter-based" enforcement.

Security needs to be an inherent feature and integral part of the architecture. Also, there is a significant demand to transform the Internet from a simple "host-to-host" packet delivery paradigm into a more diverse paradigm built around the data, content, and users instead of the machines. All of the above

challenges have led to the research on future Internet architectures.

Future Internet architecture is not a single improvement on a specific topic or goal. A cleans late solution on a specific topic may assume the other parts of the architecture to be fixed and unchanged. Thus, assembling different cleans late solutions targeting different aspects will not necessarily lead to a new Internet architecture. Instead, it has to be an overall redesign of the whole architecture, taking all the issues (security, mobility, performance reliability, etc.) into consideration. It also needs to be evolvable and flexible to accommodate future changes. Most previous clean-slate projects were focused on individual topics. Through a collaborative and comprehensive approach, the lessons learned and research results obtained from these individual efforts can be used to build a holistic Internet architecture.

V. KEY RESEARCH TOPICS

*I*n this section, we discuss some key research topics that are being addressed by different research projects.

Content- or data-oriented paradigms: Today's Internet builds around the "narrow waist" of IP, which brings the elegance of diverse design above and below IP, but also makes it hard to change the IP layer to adapt for future requirements. Since the primary usage of today's Internet has changed from host-to-host communication to content distribution, it is desirable to change the architecture's narrow waist from IP to the data or content distribution. Several research projects are based on this idea. This category of new paradigms introduces challenges in data and content security and privacy, scalability of naming and aggregation, compatibility and co-working with IP, and efficiency of the new paradigm.

Mobility and ubiquitous access to networks: The Internet is experiencing a significant shift from PC-based computing to mobile computing. Mobility has become the key driver for the future Internet. Convergence demands are increasing among heterogeneous networks such as cellular, IP, and wireless ad hoc or sensor networks that have different technical standards and business models. Putting mobility as the norm instead of an exception of the architecture potentially nurtures future Internet architecture with innovative scenarios and applications. Many collaborative research projects in academia and industry are pursuing such research topics with great interest.

These projects also face challenges such as how to trade off mobility with scalability, security, and privacy protection of mobile users, mobile endpoint resource usage optimization, and so on.

Cloud-computing-centric architectures: Migrating storage and computation into the "cloud" and creating a "computing utility" is a trend that demands new Internet services and applications. It creates new ways to provide global-scale resource provisioning in a "utilitylike" manner. Data centers are the key components of such new architectures. It is important to create secure, trustworthy, extensible, and robust architecture to interconnect data, control, and management planes of data centers. The cloud computing perspective has attracted considerable research effort and industry projects toward these goals. A major technical challenge is how to guarantee the trustworthiness of users while maintaining persistent service availability.

Security: Security was added into the original Internet as an additional overlay instead of an inherent part of the Internet architecture. Now security has become an important design goal for the future Internet architecture. The research is related to both the technical context and the economic and public policy context. From the technical aspect, it has to provide multiple granularities (encryption, authentication, authorization, etc.) for any potential use case. Also, it needs to be open and extensible to future new security related solutions. From the non-technical aspect, it should ensure a trustworthy interface among the participants (e.g., users, infrastructure providers, and content providers). There are many research projects and working groups related to security. The challenges on this topic are very diverse, and multiple participants make the issue complicated.

Experimental testbeds: As mentioned earlier, developing new Internet architectures requires large-scale testbeds. Currently, testbed research includes multiple testbeds with different virtualization technologies, and the federation and coordination among these testbeds. Research organizations from the United States, European Union, and Asia have initiated several programs related to the research and implementation of large-scale testbeds. These projects explore challenges related to large-scale hardware, software, distributed system test and maintenance, security and robustness, coordination, openness, and extensibility.

Besides these typical research topics, there are several others, including but not limited to networked multimedia; "smart dust," also called the "Internet of things"; and Internet services architecture. However, note that in this survey, we are not trying to enumerate all the possible topics and corresponding research projects. Instead, we focus on a representative subset and discuss a few important ongoing research projects. Due to length limitations, we are not able to enumerate all the references for the projects discussed below. However, we do have a longer survey [18], which includes a more complete reference list for further reading.

Since the primary usage of the today's Internet has changed from host-to-host communication to content distribution, it is desirable to change the architecture's narrow waist from IP to the data or content distribution.

Named Data Networking (NDN)

The basic argument of the NDN[3] project is that the primary usage of the current Internet has changed from end-to-end packet delivery to a content-centric model. The current Internet, which is a "client-server" model, is facing challenges in supporting secure content-oriented functionality. In this information dissemination model, the network is "transparent" and just forwarding data (i.e., it is "content-unaware"). Due to this unawareness, multiple copies of the same data are sent between endpoints on the network again and again without any traffic optimization on the network's part. The NDN uses a different model that enables the network to focus on "what" (contents) rather than "where" (addresses). The data are named instead of their location (IP addresses). Data become the first-class entities in NDN. Instead of trying to secure the transmission channel or data path

through encryption, NDN tries to secure the content by naming the data through a security-enhanced method. This approach allows separating trust in data from trust between hosts and servers, which can potentially enable content caching on the network side to optimize traffic.

NDN has several key research issues. The first one is how to find the data, or how the data are named and organized to ensure fast data lookup and delivery. The proposed idea is to name the content by a hierarchical "name tree" which is scalable and easy to retrieve. The second research issue is data security and trustworthiness. NDN proposes to secure the data directly instead of securing the data "containers" such as files, hosts, and network connections. The contents are signed by public keys. The third issue is the scaling of NDN. NDN names are longer than IP addresses, but the hierarchical structure helps the efficiency of lookup and global accessibility of the data.

Regarding these issues, NDN tries to address them along the way to resolve the challenges in routing scalability, security and trust models, fast data forwarding and delivery, content protection and privacy, and an underlying theory supporting the design.

MobilityFirst —The[4] basic motivation of MobilityFirst is that the current Internet is designed for interconnecting fixed endpoints. It fails to address the trend of dramatically increasing demands of mobile devices and services. The Internet usage and demand change is also a key driver for providing mobility from the architectural level for the future Internet. For the near term, MobilityFirst aims to address the cellular convergence trend motivated by the huge mobile population of 4 to 5 billion cellular devices; it also provides mobile peer-to-peer (P2P) and infostation (delay-tolerant network [DTN]) application services which offer robustness in case of link/network disconnection. For the long term, in the future, MobilityFirst has the ambition of connecting millions of cars via vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) modes, which involve capabilities such as location services, georouting, and reliable multicast. Ultimately, it will introduce a pervasive system to interface human beings with the physical world, and build a future Internet around people.

The challenges addressed by MobilityFirst include stronger security and trust requirements due to open wireless access, dynamic association, privacy concerns, and greater chance of network failure. MobilityFirst targets a clean-slate design directly addressing mobility such that the fixed Internet will be a special case of the general design. MobilityFirst builds the "narrow waist" of the protocol stack around several protocols: • Global name resolution and routing service

- Storage-aware (DTN-like) routing protocol
- Hop-by-hop segmented transport
- Service and management application programming interfaces (APIs)

The DTN-like routing protocol is integrated with the use of self-certifying public key addresses for inherent trustworthiness. Functionalities such as context- and location-aware services fit into the architecture naturally. It shows all the building blocks mentioned above and how they work together.

eXpressive Internet Architecture (XIA) —As[5] we observe, most of the research projects on future Internet architectures realize the importance of security and consider their architecture carefully to avoid the flaws of the

original Internet design. However, XIA directly and explicitly targets the security issue within its design.

There are three key ideas in the XIA architecture:Define a rich set of building blocks or communication entities as network principals including hosts, services, contents, and future additional entities.

- It is embedded with intrinsic security by using self-certifying identifiers for all principals for integrity and accountability properties
- A pervasive "narrow waist" (not limited to the host-based communication as in the current Internet) for all key functions, including access to principals, interaction among stakeholders, and trust management; it aims to provide interoperability at all levels in the system, not just packet forwarding.

The core of the XIA is the Expressive Internet Protocol (XIP) supporting communication between various types of principals. Three typical XIA principal types arecontent, host (defined by "who"), and service (defined by what it does). They are open to future extension. Each type of principal has a narrow waist that defines the minimal functionality required for interoperability. Principles talk to each other using expressive identifiers (XIDs), which are 160 bit identifiers identifying hosts, pieces of content, or services. The XIDs are basically self-certifying identifiers taking advantage of cryptographic hash technology. By using this XID, the content retrieval no longer relies on a particular host, service or network path. XIP can then support future functions as a diverse set of services. For low-level services, it uses a path-segment-based network architecture (named Tapa in their previous work) as the basic building block; and builds services for content-transfer and caching and service for secure content provenance at a higher level. XIA also needs various trustworthy mechanisms and provides network availability even when under attack. Finally, XIA defines explicit interfaces between network actors with different roles and goals.

GLOBAL ENVIRONMENT FOR NETWORK INNOVATIONS (GENI)

GENI [6] is a collaborative program supported by NSF aimed at providing a global large-scale experimental testbed for future Internet architecture test and validation. Started in 2005, it has attracted broad interest and participation from both academia and industry. Besides its initial support from existing projects on a dedicated backbone network infrastructure, it also aims to attract other infrastructure platforms to participate in the federation — the device control framework to provide these participating networks with users and operating environments, to observe, measure, and record the resulting experimental outcomes. So generally, GENI is different from common testbeds in that it is a generalpurpose large-scale facility that puts no limits on the network architectures, services, and applications to be evaluated; it aims to allow clean-slate designs to experiment with real users under real conditions. The key idea of GENI is to build multiple virtualized *slices* out of the substrate for resource sharing and experiments. It contains two key pieces:

- Physical network substrates that are expandable building block components

- A global control and management framework that assembles the building blocks together into a coherent facility

Thus, intuitively two kinds of activities will be involved in GENI testbeds: one is deploying a prototype testbed federating different small and medium ones together (e.g., the OpenFlow testbed for campus networks [8]); the other is to run observable, controllable, and recordable experiments on it.
There are several working groups concentrating on different areas, such as the control framework working group; GENI experiment workflow and service working group; campus/operation, management, integration, and security working group; and instrumentation and management working group.
The GENI generic control framework consists of several subsystems and corresponding basic entities: • Aggregate and components

- Clearinghouse
- Research organizations, including researchers and experiment tools
- Experiment support service
- "Opt-in" end users
- GENI operation and management

Clearinghouses from different organizations and places (e.g., those from the United States and European Union) can be connected through federation. By doing this, GENI not only federates with identical "GENI-like" systems, but also with any other system if they comply with a clearly defined and relatively narrow set of interfaces for federation. With these entities and subsystems, "slices" can be created on top of the shared substrate for miscellaneous researchdefined specific experiments, and end users can "opt in" onto the GENI testbed accordingly.
GENI's research and implementation plan consists of multiple continuous *spirals* (currently in spiral 3). Each spiral lasts for 12 months. Spiral 1 ran from 2008 to 2009; spiral 2 ran from 2009 to 2010; spiral 3 started in 2011. In spiral 1, the primary goals were to demonstrate one or more early prototypes of the GENI control framework and end-to-end slice operation across multiple technologies; there were five competing approaches to the GENI control framework,
called "clusters."
Cluster A was the Trial Integration Environment based on DETER (TIED) control framework focusing on federation, trust, and security. It was a one-project cluster based on the Cyber Defense Technology Experimental Research (DETER) control framework by the University of Southern California (USC)/ISI, which is an individual "mini-GENI" testbed to demonstrate federated and coordinated network provisioning. Cluster A particularly aimed to provide usability across multiple communities through federation. The project delivered software "fedd" as the implementation of the TIED federation architecture providing dynamic and on-demand federation, and interoperability across ProtoGENI, GENIAPI, and non-GENI aggregate. It included an Attribute Based Access Control (ABAC) mechanism for large-scale distributed systems. It created a federation with two other projects: StarBED in Japan and ProtoGENI in the United States.
Cluster B was a control framework based on PlanetLab implemented by Princeton University emphasizing experiments with virtualized machines over the Internet. By the end of spiral 2, it included at least 12 projects from different universities and research institutes. The results of these projects are to be integrated into the PlanetLab

testbed. PlanetLab provided "GENIwrapper" code for independent development of an aggregate manager (AM) for Internet entities. A special "lightweight" protocol was introduced to interface PlanetLab and OpenFlow equipment. Through these mechanisms, other projects in the cluster can design their own substrates and component managers with different capacities and features.

Cluster C was the ProtoGENI control framework by the University of Utah based on Emulab, emphasizing network control and management. By the end of spiral 2, it consisted of at least 20 projects. The cluster integrated these existing and under-construction systems to provide key GENI functions. The integration included four key components: a backbone based on Internet2; sliceable and programmable PCs and NetFPGA cards; and subnets of wireless and wired edge clusters. Cluster C so far is the largest set of integrated projects in GENI.

Cluster D was Open Resource Control Architecture (ORCA) from Duke University and RENCI focusing on resource allocation and integration of sensor networks. By the end of spiral 2, it consisted of five projects. ORCA tried to include optical resources from the existing Metro-Scale Optical Testbed (BEN). Different from other clusters, the ORCA implementation included the integration of wireless/sensor prototypes. It maintains a clearinghouse for the testbeds under the ORCA control framework through which it connects to the national backbone and is available to external researchers.

Cluster E was Open-Access Research Testbed for Next-Generation Wireless Networks (ORBIT) by Rutgers University focusing on mobile and wireless testbed networks. It included three projects by the end of spiral 2. The basic ORBIT did not include a full clearinghouse implementation. Cluster E tried to research how mobile and wireless work can affect and possibly be merged into the GENI architecture. WiMAX is one of the wireless network prototypes in this cluster.

In this article, we present a survey of the current research efforts on future Internet architectures. It is not meant to be a complete enumeration of all such projects. Instead, we focus on a series of representative research projects. Research programs and efforts from the United States, European Union, and Asia are discussed. By doing this, we hope to draw an approximate overall picture of the up-to-date status in this area.

Acknowledgements

Hereby we would like to thank Ms. Kanika Jethwani for their help, right guidance and motivation which they provided us right from the beginning to the end.

REFERENCES

[1] NSF Future Internet Architecture Project, http://www. nets-fia.net/.

[2] Named Data Networking Project, http://www.nameddata.net.

[3] MobilityFirst Future Internet Architecture Project, http://mobilityfirst.winlab.rutgers.edu/.

[4] eXpressive Internet Architecture Project, http://www.cs.cmu.edu/~xia/.

[5] Global Environment for Network Innovations (GENI) Project, http://www.geni.net/

MALWARE ANALYSIS AND INTRUSION DETECTION

JATIN ARORA, KAMAL SINGH RAWAT, SHUBHI BANSAL
Institute of Innovation in Technology & Management
New Delhi
arora69.bunny@gmail.com
ksr00710@gmail.com
shubhi.bansal67@gmail.com

KEYWORDS
Firewall- Barrier between untrusted and trusted networks,
Server - Server is a computer program or a device that provides functionality for other programs or devices, called clients.
Switch- Device for making and breaking the connection in an electric circuit.
Worms- It is a self-replicating virus that does not alter files but resides in active memory and duplicates itself.
Trojan Horse- A person or thing intended to undermine or secretly overthrow an enemy or opponent.
Virus- A **virus** is a small infectious agent that replicates only inside the living cells of other organisms.

1.INTRODUCTION

Now a day, Internet has become an essential part of the daily life of many people. On Internet many services are available and also increasing day by day. More and more people are making use of these services. The Internet has evolved from a basic communication network to an interconnected set of information sources enabling, among other things, new forms of interactions and market places for the sale of products and services. Online banking or advertising are the examples of the commercial services of the Internet. Just as in the physical world, there are people on the Internet with malevolent intents that strive to enrich themselves by taking advantage of legitimate users whenever money is involved. Malware like software of malicious intent helps these people accomplishing their goals. There are two types of malware analysis, Static and Dynamic.

Malware analysis is the study or process of determining the functionality, origin and potential impact of a given malware sample such as a virus, worm, Trojan horse, rootkit, or backdoor. Malware or malicious software is any computer software intended to harm the host operating system or to steal sensitive data from users, organizations or companies. Malware may include software that gathers user information without permission.

Malware analysis is the study or process of determining the functionality, origin and potential impact of a given malware sample such as a virus, worm, trojan horse.

Intrusion Detection System (IDS) is meant to be a software application which monitors the network or system activities and finds if any malicious operations occur. Tremendous growth and usage of internet raises concerns about how to protect and communicate the digital information in a safe manner. Nowadays, hackers use different types of attacks for getting the valuable information. Many intrusion detection techniques, methods and algorithms help to detect these attacks. This main objective of this paper is to provide a complete study about the definition of intrusion detection, history, life cycle, types of intrusion detection

methods, types of attacks, different tools and techniques, research needs, challenges and applications.

The purpose of this report is to introduce the user to Intrusion Detect Systems and give a deep understanding of some sophisticated techniques for intrusion detection. Intrusion Detection is an important component of infrastructure protection mechanisms. Given the increasing complexities of today's network environments, more and more hosts are becoming vulnerable to attacks and hence it is important to look at systematic, efficient and automated approaches for Intrusion Detection. Here we discuss some data mining based approaches for intrusion detection and compare their merits and demerits. We also look at some signature based detection techniques for detecting polymorphic worms. We also look at various port scanning techniques and discuss some techniques for detecting port scanning attempts. We then discuss the architecture of an advance Intrusion Detection System, Snort and suggest some enhancements to the same.

MALWARE ANALYSIS

What is Malware?

Malware stands for malicious software, designed to damage or to infiltrate a computer system without the owner's informed consent. Viruses, Worms, Trojan, Keyloggers and Spyware are the examples of malware. In other words we can also say Software that "deliberately fulfils the harmful intent of an attacker" is commonly referred to as malicious software or malware. Terms, such as "worm", "virus", or "Trojan horse" are used to classify malware samples that exhibits similar behavior.

What is Malware Analysis?

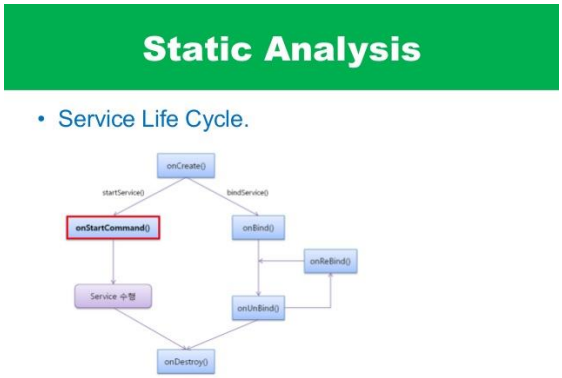
Malware analysis is the process of determining the purpose and functionality of a given malware sample such as a virus, worm, or Trojan horse. This process is a necessary step to be able to develop effective detection techniques for malicious code. In addition, it is an important prerequisite for the development of removal tools that can thoroughly delete malware from an infected machine. Traditionally, malware analysis has been a manual process that is tedious and time-intensive. Unfortunately, the number of samples that need to be analyzed by security vendors on a daily basis is constantly increasing. The process of analyzing a given program during execution is called dynamic analysis; while static analysis refers to all techniques that analyze a program by inspecting it.

Two types of Malware Analysis

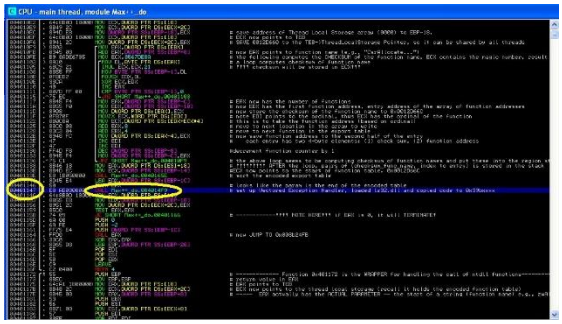
- Static Malware Analysis
- Dynamic Malware Analysis

STATIC MALWARE ANALYSIS

Analyzing software without executing it is called static analysis. Static analysis techniques can be applied on different representations of a program. If the source code is available, static analysis tools can help finding memory corruption flaws and prove the correctness of models for a given system Static analysis tools can also be used on the binary representation of a program. When compiling the source code of a program into a binary executable, some information gets lost. This loss of information further complicates the task of analyzing the code.

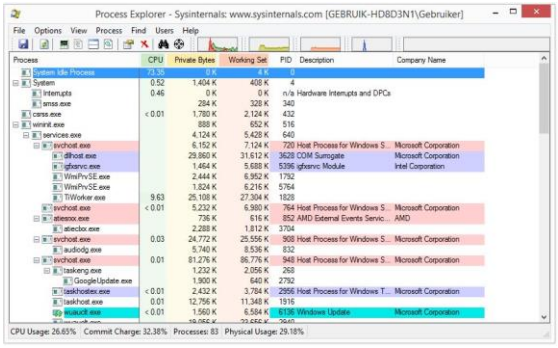


The process of inspecting a given binary without executing it is mostly conducted manually. For example, if the source code is available several interesting information, such as data structures, used functions and call graphs can be extracted. This information gets lost once the source code has been compiled into a binary executable and thus impedes further analysis. Within the malware domain typically the latter is the case, since the source code of a current malware binary is typically not available

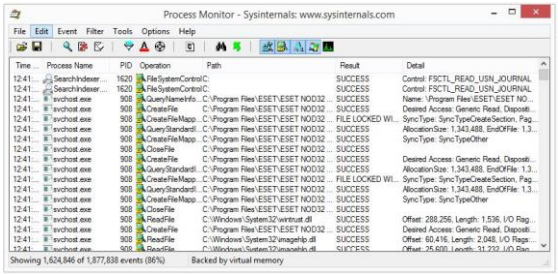


DYNAMIC MALWARE ANALYSIS

Executing a given malware sample within a controlled environment and monitoring its actions in order to analyze the malicious behavior is called dynamic malware analysis. Since Dynamic Malware Analysis is performed during runtime and malware unpacks itself, dynamic malware analysis evades the restrictions of static analysis. Thereby it is easy to see the actual behavior of a program. Another major advantage is that it can be automated thus enabling analysis at a large scale basis. However, the main drawback is so-called dormant code: That is, unlike static analysis, dynamic analysis usually monitors only one execution path and thus suffers from incomplete code coverage. In addition there is the danger of harming third party systems, if the analysis environment is not properly isolated or restricted respectively. Furthermore, malware samples may alter their behavior or stop executing at all once they detect to be executed within a controlled analysis environment.



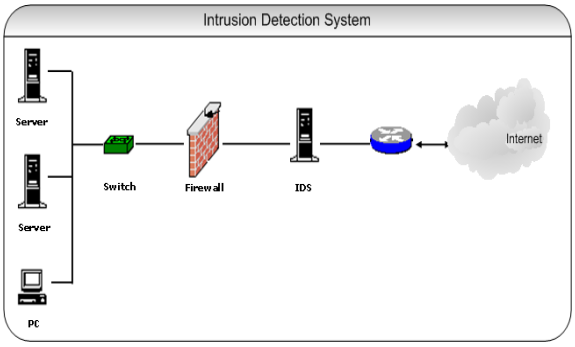
Mainly two basic approaches for dynamic malware analysis can be distinguished: • Analyzing the difference between defined points: A given malware sample is executed for a certain period of time and afterwards the modifications made to the system are analyzed by comparison to the initial system state. In this approach, Comparison report states behavior of malware. Observing runtime-behavior: In this approach, malicious activities launched by the malicious application are monitored during runtime using a specialized tool.



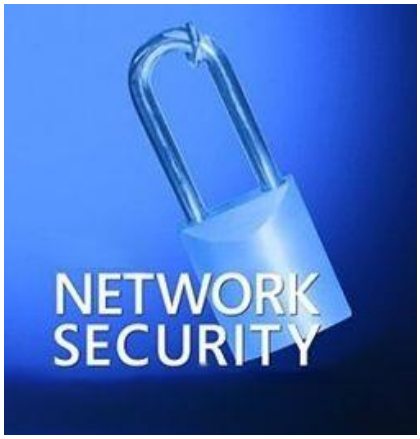
INTRUSION DETECTION

An Intrusion Detection System is an application used for monitoring the network and protecting it from the intruder. With the rapid progress in the internet based technology new application areas for computer network have emerged. In instances, the fields like business, financial, industry, security and healthcare sectors the LAN and WAN applications have progressed. All of these application areas made the network an attractive target for the abuse and a big vulnerability for the community. Malicious users or hackers use the organization's internal systems to collect information's and cause vulnerabilities like Software bugs, Lapse in administration, leaving systems to default configuration. As the internet emerging into the society, new stuffs like viruses and worms are imported. The malignant so, the users use different techniques like cracking of password, detecting unencrypted text are used to cause vulnerabilities to the system. Firewall technique is one of the popular protection techniques and it is used to protect the private network from the public network. IDS are used in network related activities, medical applications, credit card frauds, Insurance agency.

An IDS is composed of the following three components: Sensors: - which sense the network traffic or system activity and generate events. Console: - to monitor events and alerts and control the sensors, Detection Engine: - that records events logged by the sensors in a database and uses a system of rules to generate alerts from the received security events.



An IDS is referred as burglar alarm. For example the lock system in the house protects the house from theft. But if somebody breaks the lock system and tries to enter into the house, it is the burglar alarm that detects that the lock has been broken and alerts the owner by raising an alarm. Moreover, Firewalls do a very good job of filtering the incoming traffic from the Internet to circumvent the firewall. For example, external users can connect to the Intranet by dialing through a modem installed in the private network of the organization; this kind of access cannot be detected by the firewall.
There are two types of prevention system they are Network (NIPS) and Host (HIPS). These systems watch the network traffic and automatically take actions to protect networks and systems. IPS issue is false positives and negatives. False positive is defined to be an event which produces an alarm in IDS where there is no attack. False negative is defined to be an event which does not produces an alarm when there is an attacks takes place. Inline operation can create bottlenecks such as single point of failure, signature updates and encrypted traffic. The actions occurring in a system or network is measured by IDS.



- In terms of performance, an intrusion detection system becomes more accurate as it detects more attacks and raises fewer false alarms.

REFERENCES

- http://mashable.com/2011/02/03/ipv4-ipv6-guide/#kjB_TPYLuOqr
- http://www.webopedia.com
- http://www.linksys.com/in/
- https://en.wikipedia.org/

CONCLUSION

- Dynamic malware analysis is the best way to analyze malware samples.
- Intrusion detection systems search for signs of an attack and flag when an intrusion is detected

Complex Procedures Carried out by Robots in Surgery

Lakshay Sharma
lakshaysharma096@gmail.com
IINTM , Janakpuri , New Delhi

Harshit Batra
harshitbatra18@gmail.com
IINTM, Janakpuri ,New Delhi

Shiv Kumar
shivkm1996@gmail.com
IINTM,Janakpuri,New Delhi

Abstract **- The Robotic Surgery anonymous Robot-assisted Surgery is about surgery with the use of robots instead of processing various major operations manually. In earlier time, operations are taking place manually due to which there were risk and according to the earlier doctors, the surgery is about the cutting and sewing with hand instruments, direct visualization and direct contact with organs. The robotics surgery is the technique in which doctor will do the operation but by the use of the computers and robots, which promises to facilitate the complex procedures by the use of voice control, the networked**

operating room and the ability to enhance the ability to learn new complex operations. Doctors of 21st had started performing complex operations by the use of Robotics instruments, which give them chances for more success. The use of Robotics instrument not only deal with complex operation but also with the minor operations or therapy which require more focusing on the certain part of the body .In future, the research will be more focused on delivery of diagnostic and therapeutic modularity through natural opening in which investigation is under the remote control and navigation.

Keyword—Robots, Network operating room, Surgery, Therapy, Remote control

I. Introduction

The use of robots taken place in many fields , but more focus was taken place in medical field. The robots initially used in cartoons as a material to kids but then it came in reality in 1991 and called as probot which stands for prostate robots. This was the first robot used in medical field and a special-purpose robot had been devised and clinically applied to independently remove tissue from a human patient. Since that time there have been many robotic surgery research projects and a small number which resulted in companies who have produced systems that have been applied clinically. However it is surprising that relatively few robotic procedures have been undertaken clinically. It is this aspect that will be the focus of this paper and an attempt made to suggest reasons and how we might best proceed in the future. Robotic surgery is a new and emerging technology that is useful for numerous types of surgeries. Controversy between robot manufacturers and surgeons continuously challenge its operational feasibility and legitimacy. This research supported essay will provide in-depth information in order to increase awareness and understanding about robotic surgery. Additionally, areas such as legal and ethical issues, social consequences, and security concerns will be addressed. As a result of reading, the reader will receive a wealth of information regarding the world changing phenomena that is, robotic surgery.[1]

In 1928, one of the first humanoid robots was exhibited at the annual exhibition of the Model Engineers Society in London. Invented by W. H. Richards, the robot Eric's frame consisted of an aluminum body of amour with eleven electromagnets and one motor powered by a twelve-volt power source. The robot could move its hands and head and could be controlled through remote control or voice control. By the use of such Technology an invent of Surgery taken place in which robots are used. Such a technology is called as Robotics Surgery

Fig 1:This the use of the robots that are being present in the operation theatre and the doctor are operating the patient from outside

II. BENEFITS OF ROBOTICS SURGERY

The introduction to Surgery through Robots donot only facilitate the enhancement of Medical field but it had various benefits to all people associated to it.

- They bring down the cost of the operation.
- The patient has a faster recovery time because there is fewer traumas.
- There is less chance of complication.
- The doctor can see better through fiber optics than with his or her regular sight.
- Robotic arms do not suffer fatigue or tremors as do the human hand.

III. ROBOTIC SYSTEM REQUIREMENTS

Bandwidth *:* This category of application, command and control usually requires relatively low bandwidth. This appears to be the case for the control of medical devices as well. The exception is the robotic surgery application which will require not just the transmission of control signals, but also the transmission of real-time motion images at VMS tape quality (about 75 Mbits/second).

Latency: Latency is the essential factor in these applications as the overshoot of a controlled parameter could result in a life-threatening situation. It has been suggested that the latency of the human nervous system is under 15 msec, and therefore the latency of the control system should be under 10 msec.

Security *:* These applications may involve the transmission of real patient data which the patient may consider sensitive or control signal which, if tampered with, could cause life threatening situations. They require a high level of transmission security to be sure that they cannot be viewed or altered during transmission.

Reliability: The network must be as close to 100% reliable as possible as lack of reliability could result in life threatening situations.

Scalability: This group of applications tend to be point to point applications and so will initially not require scalability. However, if viewed as a successful telemedicine applications, the bandwidth available on the network should be scalable so that the capacity can be increased in the future in response to potential increases in demand. In the case of robotic surgery, future additional interactive monitoring sites may be added so

that such tele-surgery can also serve a useful teaching function which will require a degree of network scalability.[2]

How Robotic Surgery differ from Manual Surgery

The Laparoscopic surgery—in which instruments are inserted through small incisions—has been used by surgeons whenever possible. Patients are less traumatized, require shorter hospital stays, and heal faster than with conventional surgery. Laparoscopic instruments are mainly limited to scissors and staplers to close incisions or attach blood vessels. It also has graspers to manipulate tissue. The instruments enter the body through a long tube; a video image from a tiny camera called an endoscope poked through another incision guides the surgeon. For a relatively simple procedure like gallbladder removal, the tools work well enough. But surgeons can't use the instruments to perform complicated tasks like suturing and knot tying. Because of these limitations, most operations can't be performed endoscopically.

Robotic surgery uses laparoscopic tools—including miniature robotic hands with the dexterity to tie knots. The reason the surgeons have to cut a person open is to get their hands in there. The surgeons like to get their hands around the organs, to palpate them.

Robotic surgery provides with little instruments in there that let the surgeons feel as if they are working with their hands in a normal procedure, and hence avoiding a bigger incision.

The Robotic surgical system consists of a pencil-size joystick (one each for the surgeon's right and left hands), a computer, and right-hand and left-hand end effectors —the robotic instruments that snake into the body to perform the actual surgery.

Each hydraulically powered end effector consists of a single digit, three to four inches long and less than half an inch wide. It has four joints that rotate, swivel, and swing back and forth and a grasper at the end. The result: A finger that functions like an entire hand. The surgery is completely anthropomorphic. If the hand moves in, the instrument moves in; if the hand moves to the right, the instrument moves to the right.

The system also has force feedback, which relays to doctors the response of muscles and other tissues to their actions. The feedback makes the procedure feel more like normal surgery. The system also has tactile sensors that will transmit the feel of tissue to the surgeon's fingertips.[3]

Fig 2: The figure shows that involvement of human during the surgery through robots. This show that the robots cannot do operations individually.

Laparoscopic Limitations/Robotic Solutions

Laparoscopic Problems/Limitations	
Two-dimensional vision of surgical field displayed on the monitor impairs depth perception	Binocular syste
Movements are counterintuitive (ie, moving the instrument to the right appears to the left on the screen due to mirror-image effect)	Movements are on the viewer)
Unstable camera held by an assistant	Surgeon contro
Diminished degrees of freedom of straight laparoscopic instruments	Microwrists nea
Surgeon forced to adopt uncomfortable postures during operation	Superior operat
Steep learning curve	Shorter learning

Table 1 : Solutions provided by Robotics Surgery to the Laparoscopic Surgery

IV. THE FUTURE OF ROBOTIC SURGERY

The future of robotic surgery is hard to believe but ....it is now. If you haven't noticed, robotic surgery has come long ways and it was only a dream for doctors and engineers to have something that you no longer had to make big, hideous scars that would mess up somebody's body for the rest of their lives. Doctors, before robotic surgery, worked on making minimally invasive surgery that would take hours of surgery time. Now, surgery is still made in hours, but shorter hours are now in check with the robotic surgery. So you can't really say there is a future of robotic surgery, but you can say that this has been the future for doctors long ago
so all you can say is...the future is now!![4]
The Future is now the field of surgery is entering a time of great change, spurred on by remarkable recent advances in surgical and computer technology. Computer controlled diagnostic instruments have been used in the operating room for years to help provide vital information through ultrasound, computer-aided tomography (CAT), and other imaging technologies. Only recently have robotic systems made their way into the operating room as dexterity-enhancing surgical assistants and surgical planners, in answer to surgeons' demands for ways to overcome the surgical limitations of minimally invasive laparoscopic surgery, a technique developed in the 1980s.

V. Classification of the Surgical Robots

The first generation of surgical robots are already being installed in a number of operating rooms around the world. These aren't true autonomous robots that can perform surgical tasks on their own, but they are lending a mechanical helping hand to surgeons. These machines still require a human surgeon to operate them and input instructions. Remote control and voice activation are the methods by which these surgical robots are controlled. Robotics are being introduced to medicine because they allow for unprecedented control and precision of surgical instruments in minimally invasive procedures. So far, these machines have been used to position an endoscope, perform gallbladder surgery and correct gastroesophogeal reflux and heartburn. The ultimate goal of the robotic surgery field is to design a robot that can be used to perform closed-chest, beating-heart surgery. According to one manufacturer, robotic devices could be used in more than 3.5 million medical procedures per year in the United States alone. Here are three surgical robots that have been recently developed:

- Da Vinci Surgical System
- ZEUS Robotic Surgical System
- AESOP Robotics Surgical System

Da Vinci Surgical System

The Da Vinci Surgical System enables surgeons to perform operations through a few small incisions and features several key features,including:

1) Magnified vision system that gives surgeons a 3D HD view inside the patient body
2) A console which lead the doctors to perform operation easily.
3) Patient cart side where the patient is positioned during surgery.

4) Wristed instruments that bend and rotate far greater that the human hand.

Fig 3: This show Da Vinci Surgical System which arise after ZEUS Surgical System .

This system is powered by robotic technology that allows the surgeon's hand movements to be translated into smaller , precise movements of tiny instruments inside the patients body.
One of the instruments is laparoscope – a thin tube with a tiny camera and light at the end. The camera send the image of the body parts from the inner view and let the doctors to operate. [5]

ZEUS Robotic Surgical System

The **ZEUS Robotic Surgical System** (ZRSS) was a medical robot designed to assist in surgery, originally produced by the American robotics company Computer Motion. Its predecessor, AESOP, was cleared by the Food and Drug Administration in 1994 to assist surgeons in minimally invasive surgery. The ZRSS itself was cleared by the FDA seven years later, in 2001. ZEUS had three robotic arms, which were remotely controlled by the surgeon. The first arm, AESOP (Automated Endoscopic System for Optimal Positioning), was a voice-activated endoscope, allowing the surgeon to see inside the patient's body. The other two robotic arms mimicked the surgeon's movements to make precise incisions and extractions.[1] ZEUS was discontinued in 2003, following the merger of Computer Motion with its rival Intuitive Surgical; the merged company instead developed the Da Vinci Surgical System. [6]

Fig 4: The ZEUS Surgical System . The robots which was used in the surgery.

AESOP ROBOTIC SURGICAL SYSTEM

The AESOP system employs the assistance of the Automated Endoscopic System for optical position. AESOP was the first robot to be cleared by FDA for assisting surgery in the operating room. AESOP is much simpler than the da Vinci and Zeus system. It is used by the physician to position the endoscope of a surgical camera inserted into the patient. Voice activated software allow the physician

to position the camera leaving her hands free. The AESOP robotic surgical system was very complex. So that it cannot be used in operating rooms [2]

Graph for the timeline use of robots in surgery

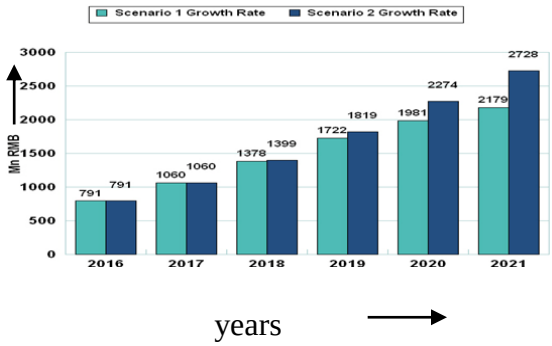


Fig 5: The Number of estimated operations to be taken place in china till 2021 [7]

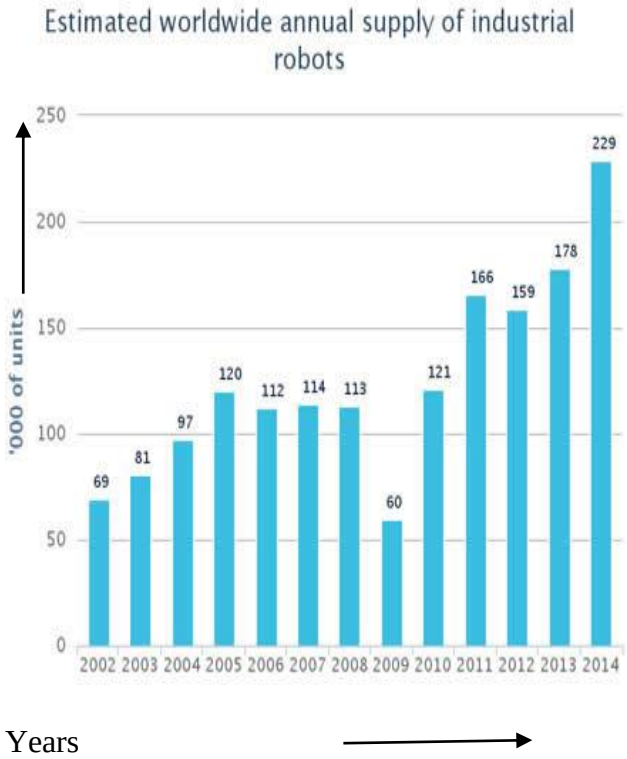


Fig 6: Number of estimated robots sold for the operations purpose[7]

The above Fig 1 and Fig 2 are correlated as they are on the term “robots” . The Fig 1 shows that the number and the estimate number of operations that can take place till the year 2021 whereas the Fig 2 shows the number of robots that are sold all over world till year 2014.

VI.APPLICATIONS

Robotic procedures spanning the whole spectrum of surgery have been successfully executed Initial results show that mortality, morbidity, and hospital stay compare favorably to conventional laparoscopic operations. However, only a limited number of randomized, prospective studies that compare outcomes of robotic techniques with conventional methods exist. More procedure-specific, randomized trials need to be performed before robotic surgery can find its way into everyday surgical practice. [7]

Field	Operations Performed via Robotic Surgery
Robotic gastrointestinal surgery	first robotic cholecystectomy
	Antireflux operations, Heller's myotomy, gastric bypass, gastrojejunostomy, esophojectomy, gastric banding colectomy, splenectomy, adrenalectomy, and pancreatic resection reported to date
Robotic urologic surgery	Radical robotic prostatectomy is the most common operation performed robotically and is gaining widespread recognition in the United States and EuropeNephrectomy and pelvic lymph node dissection also reported
Robotic gynecologic surgery	Robotic hysterectomy, salpingo-oophorectomy, and microsugical fallopian tube reanastomosis
Robotic cardiothoracic surgery	Surgical robots allow cardiothoracic surgeons to perform complex cardiothoracic procedures while avoiding the significant morbidity of sternotomy and thoracotomyHundreds of robotic coronary bypasses have been performed to dateMitral valve repairs, atrial spetal defect repair, pericardiectomy, lobectomies, and tumor enucleations
Robotic oncologic surgery	Esophageal tumors, gastric cancer, colon cancer, thymoma, and retromediastinal tumors
Robotic pediatric surgery	Pyeloplasty for ureteropelvic junction obstruction, antireflux procedures for gastroesophageal reflux disease, and pediatric congenital heart diseases, such as ligation of patent ductus arteriosus

Table 2: The above tables shows the field in which various operations are performed via Robotic Surgery

VII.ADVANTAGES

The advantages of these systems are many because they overcome many of the obstacles of laparoscopic surgery .They increase dexterity, restore proper hand-

eye coordination and an ergonomic position, and improve visualization .In addition, these systems make surgeries that were technically difficult or unfeasible previously, now possible.

These robotic systems enhance dexterity in several ways. Instruments with increased degrees of freedom greatly enhance the surgeon's ability to manipulate instruments and thus the tissues. These systems are designed so that the surgeons' tremor can be compensated on the end-effector motion through appropriate hardware and software filters. In addition, these systems can scale movements so that large movements of the control grips can be transformed into micromotions inside the patient.

Another important advantage is the restoration of proper hand-eye coordination and an ergonomic position. These robotic systems eliminate the fulcrum effect, making instrument manipulation more intuitive. With the surgeon sitting at a remote, ergonomically designed workstation, current systems also eliminate the need to twist and turn in awkward positions to move the instruments and visualize the monitor. [8]

VIII.LIMITATIONS

Although rapidly developing, robotic surgical technology has not achieved its full potential owing to a few limitations. Cost-effectiveness is a major issue 2 recent studies comparing robotic procedures with conventional operations showed that although the absolute cost for robotic operations was higher, the major part of the increased cost was attributed to the initial cost of purchasing the robot (estimated at $1,200,000) and yearly maintenance ($100,000). Both factors are expected to decrease as robotic systems gain more widespread acceptance. However, it is conceivable that further technical advances may at first drive prices even higher. Decreasing operative time and hospital stay will also contribute to the cost-effectiveness of robotic surgery. [8]

IX.CONCLUSION

Though Various obstacles had arouse in the field of the robotics but it is the one of the most eminent development in the field of the medical as it is acting as the eye opening development. Various investigation of the robotics surgery are going from creating various improvement in this field. As a result, it will never cease to elevate additional legal, ethical, social, and security concerns in the medical field. After all, it is up to research scientists and physicians to determine if the benefits outweigh the cost, and only time will tell if every operating room will utilize robotic surgery.

ACKNOWLEDGMENT

We have taken efforts in this report. However, it would not have been possible without the kind support and help of many individuals and organizations. We would like to extend my sincere thanks to all of them.

We are highly indebted to Ms. Kanika Jethwani and Mr. Ajay Phogat for their guidance and constant supervision as well as for providing necessary information regarding the report & also for their support in completing the report.

We would like to express my special gratitude and thanks to College persons for giving me such attention and time.

My thanks and appreciations also go to my guide in developing the report and people who have willingly helped me out with their abilities.

REFERENCES

[1]https://www.ecri.org/Resources/ASG/Robotic_Surgery_Infographic_MS15369_web.pdf

[2] http://dspace.cusat.ac.in/jspui/bitstream/123456789/2232/1/Robotics%20in%20Surgery.pdf

[3] Krovi_SurgRob_SpIssue_DSC_Magazine_September2015.pdf

[4] C. Cote, D. Letourneau, F. Michaud, J.-M. Valin, Y. Brousseau, C. Raiev- sky, M. Lemay,

V. Tran: Code Reusability Tools for Programming Mobile Robots, Proceedings of the 2004

IEEE/RSJ International Conference on Intellige Robots and Systems (IROS'04), pages 1820–

1825, Senda, Japan, 2004.

[5]http://www.davincisurgery.com/da-vinci-surgery/da-vinci-surgical-system/

[6] https://en.wikipedia.org/wiki/ZEUS_robotic_surgical_system

[7] https://www.google.co.in/search?espv=2&biw=1280&bih=590&tbm=isch&sa=1&q=graph+for+last++years+that+the+number+of+person+are+operated+through+robotics+surgery&oq=graph+for+last++years+that+the+number+of+person+are+operated+through+robotics+surgery&gs_l=img.3...9372.9372.0.10691.1.1.0.0.0.0.186.186.0j1.1.0....0...1c.1.64.img..0.0.0.Tjw6gA4f8bA#imgrc=pbw93eUuAId8yM:

[8]https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1681689/(Table 1 & 2)

[9]http://www.medscape.com/viewarticle/466691_4

UNIVERSAL MOBILE TELECOMMUNICATION SYSTEMS

Smriti Nagrath (*Author*)
BCA (07190302014)
Institute of Innovation in Technology and Management
Guru Gobind Singh Indraprastha University
New Delhi , India
Email: smritinagrath7@gmail.com

Priya Singh (*Author*)
BCA (07490302014)
Institute of Innovation in Technology and Management
Guru Gobind Singh Indraprastha University
New Delhi , India
Email: priya.singh1042@gmail.com

Priya Chaudhary(*Author)*
BCA(07090302014)
Institute of Innovation in Technology and Management
Guru Gobind Singh University
New Delhi, India
Email: p1995chaudhary@gmail.com

Abstract— A lot has been written about 3G networks in the last few years . After a lot of hype and frustration these networks are currently deployed in many countries around the world. The International Telecommunication Union (ITU) made a request for proposals for radio transmission technology(RTT) for the international mobile telecommunication (IMT) 2000 program . IMT-2000 , formerly called future public land mobile telecommunication system (FPLMT) , tried to establish a common worldwide communication system.

The European proposal for IMT-2000 prepared by ETSI is called Universal Mobile Telecommunication System (UMTS) ,the specific proposal for radio interface RTT is UMTS. UMTS (Universal Mobile Telecommunications System) is a so-called "third-generation (3G)," broadband , packet -based transmission of text, digitized voice, video, and multimedia at data rates up to and possibly higher than 2 megabits per second (Mbps), offering a consistent set of services to mobile computer and phone users no matter where they are located in the world. Based on the Global System for Mobile (GSM) communication standard, UMTS, endorsed by major standards bodies and manufacturers, is the planned standard for mobile users around the world by 2002. Once UMTS is fully implemented, computer and phone users can be constantly attached to the Internet as they travel and, as they roaming service , have the same set of capabilities no matter where they travel to.

Keywords- GSM (Global System for Mobile Communication), UMTS, 3G (Third Generation) , roaming service , packet- based

Introduction

"Universal Mobile Telecommunications System", UMTS is represented as an evolution in terms of capacity, data speed and new service capabilities from second generation mobile networks. Today, more than 60 UMTS networks are using WCDMA technology and are operating commercially in 25

countries, supported by 100 terminal designs from Asian, European and US manufacturers.

A key member of the global family of third generation (3G) mobile technologies identified by the ITU, 3G offers mobile operators a significant capacity and broadband capabilities to support greater numbers of voice and data customers - especially in urban centers and provide higher data rates at lower incremental cost than 2G. The use of radio spectrum in bands was identified by the ITU for Third Generation IMT-2000 mobile services and subsequently licensed to operators, UMTS employs a 5 MHz channel carrier width to deliver significantly higher data rates and increased capacity compared with 2G networks. This 5 MHz channel carrier provides optimum use of radio resources, especially for operators who have been granted large, contiguous blocks of

spectrum - typically ranging from 2x10 MHz up to 2x20 MHz .This was done to reduce the cost of deploying 3G networks.

UMTS Architecture

The UMTS 3G architecture is required to provide a greater level of performance to that of the original GSM network. However as many networks had migrated through the use of GPRS and EDGE, they already had the ability to carry data. Accordingly many of the elements required for the WCDMA / UMTS network architecture were seen as a migration. This considerably reduced the cost of implementing the UMTS network as many elements were in place or needed upgrading.

With one of the major aims of UMTS being to be able to carry data, the UMTS network architecture was designed to enable a considerable improvement in data performance over that provided for GSM.

Figure 1 shows the very simplified UMTS reference architecture which applies to both UTRA solutions (3GPP, 2000). The **UTRA network (UTRAN)** handles cell level mobility and comprises of several **radio network subsystems (RNS)**. The functions of RNS includes radio channel ciphering and deciphering, hand-over control, radio resource management etc. The UTAN is connected to the **user equipment (UE)** via radio interface $\mathbf{U_u}$ (which is comparable to the U_m interface in GSM). Via the $\mathbf{I_u}$ interface (which is similar to A interface in GSM) , UTRAN communicates with the **core network (CN).** The CN contains functions of inter-

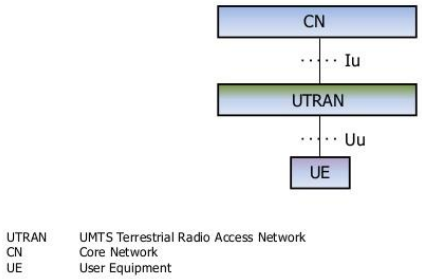


Figure.1

system handover, gateways to other networks (fixed or wireless) , and performs location management if there is no dedicated connection between UE and UTRAN.

UMTS further subdivides the above simplified architecture into so-called **domains (figure2).** The user equipment domain is assigned to the single user and comprises of all the functions that are needed to access UMTS services. Within this domain the SIM for UMTS which performs functions for encryption and authentication of users , and stores all the necessary user-related data for UMTS. Typically , this UMTS belongs to the service provider and contains a micro processor for an enhanced program execution environment(USAT, UMTS SIM application toolkit). The end device itself is in the mobile equipment domain . All functions for radio

transmission as well as user interfaces are located here.

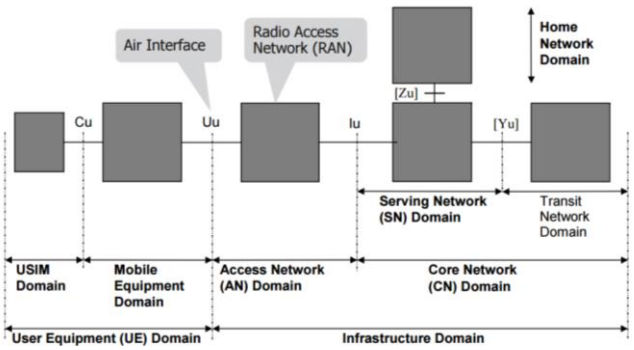


Figure.2

The infrastructure domain is shared among all users and offers UMTS services to all accepted users. This domain consists of the **access network** domain, which contains the radio access networks (RAN) , and the **core network** domain which contains access network independent functions . The core network domain can be separated into three domains with specific tasks. The serving network domain comprises all functions currently used by a user for accessing UMTS services. All the functions related to the home network of a user , e.g, user data look-up , fall into the **home network domain** . Finally , the **transit network** domain may be necessary if , for example, the serving network cannot directly contact the home network . All three domains within core network may be in fact the same physical network.

UMTS releases and standardization

The initial release of the UMTS standards was called **release 99 or R99** for short. This release of the specification describes the new radio access technologies UTRA FDD and UTRA TDD , and standardizes the use of a GSM / GPRS network as core within 440 separate specifications. This enables the cost effective migration from GSM to UTMS . the initial installation will even offer the FDD mode only as indicated in Figure3. This release was (almost) finalized in 1999 –hence the R99.

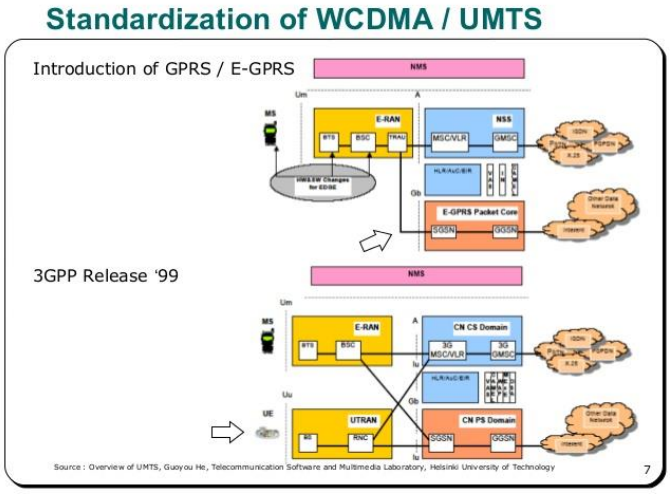


Figure.3

After R99 the release 200n0 or R00 followed. However , in September 2000 3GPP realized that it would impossible to finalize the standard within the year 2000. 3GPP decided to split R2000 into two standards and call them release 4 (Rel-4)and release 5(Rel-5). The version of all standards finalized for R99 start with 3.x.y(a reason for renaming R99 into Rel-3), Rel-4 and Rel-5 versions start with 4.x.y and 5.x.y, respectively. The standards are grouped into series.

Release 4- introduces quality of service in the fixed network plus several executions environments (example , MExE, mobile execution environment) and new service architecture. Furthermore , the Chinese proposal , TD-SCDMA was added as low chip rate option to UTRA-TDD. This releasw already consists of over 500 specifications and was frozen in March 2001.

Release 5-specifies the radically different core network. The GSM/GPRS based network will be replaced by an almost all-IP-core. While the radio interfaces remain the same , the changes in the core are tremendous for telecommunication network operators who have used traditional telephone technologies for many years. This was frozen in March 2002. This standardization integrates IP-based multimedia services(IMS)controlled by the IETF's session initiated protocol(SIP, TFC3261; Rosenberg , 2002 ;SIP Forum , 2002). A high speed downlink packet access (HSDPA) with speeds in the order of 8-10Mbit/s was added as well as a wideband 16khz AMR codec for better audio quality. Additional featured are end to end QoS messaging and several data compression mechanisms .

Release 6-3GPP is currently working on release 6 which is expected to be frozen in march 2003. This release comprises the use of multiple inputs and multiple outputs (MIMO) antennas , enhances MMS, security enhancement, WLAN/UMTS internetworking , broadcast/ multicast services, enhanced IMS, IP emergency calls , and many more management features (3GPP, 2002a).

UMTS vs. GSM

GSM stands for Global system for mobile evolution, developed based on 3GPP standards. GSM network is consists of Mobile station, Base

station subsystem(BTS, BSC) and Network(MSC) and operation subsystem. In GSM, data is supported based on GPRS technology.

UMTS stands for Universal Mobile Telecommunications System, developed based on 3GPP standards. UMTS network composed of three main parts UE (User Equipment),Radio Access Network (RAN) and Core Network various technologies fall under UMTS based on different releases from 3GPP community. UMTS is also referred as 3G. WCDMA was introduced in R99. Later HSDPA,HSUPA,HSPA,LTE and LTE advanced were introduced consecutive releases from 3GPP. Though UMTS supports both voice and data similar to GSM, certain standards for example HSPA was targeted mainly for increasing the data rate capabilities of UMTS mobile terminals/dongles.

Table.1

Features	**GSM**	**UMTS**
Frequency Band(MHz)	There are various bands in GSM, the major among them are 850MHz, 900MHz, 1800MHz and 1900MHz,	There are various bands in UMTS from Band-I to Band-VI, Each band specifies frequency and UARFCN
Carrier Spacing(KHz)	200 KHz	1230 KHz (CDMA version)
Multiple Access technique	FDMA/TDMA	CDMA
Frame	Frame duration is about 4.615ms	Frame duration is about 10ms (consisting of 15 slots)
Channel rate (Kb/s)	270.833 Kbps	1228.8 kbps (in CDMA)
Standard versions	There are various releases such as R99(WCDMA), R5 (HSDPA), R6(HSUPA) etc.	There are various documents specified by ETSI and 3GPP, 3GPP™TS 45-series

1 Early version of W-CDMA specified a clipping rate of 4.096 M chip/s and 16 times slot per frame. This was changed during the harmonization process which was necessary to avoid patent conflicts and to enable devices that can handle different CDMA standards. The harmonization process if fostered by the operators harmonized group(OHG). OHG was founded in 1999.

UTRA-FDD(W-CDMA)

The FDD mode for UTRA uses wideband CDMA (W-CDMA)with direct sequence spreading. As implied by FDD , uplink and downlink use different frequencies. A mobile station in Europe sends via the uplink using a carrier between 1920 and 1980MHz, the base station uses 2110 to 2170MHzfor the downlink. Figure 4 shows a radio frame comprising 15 times slot. Time slots in W_CDMA are not used for separation but to support periodic functions.

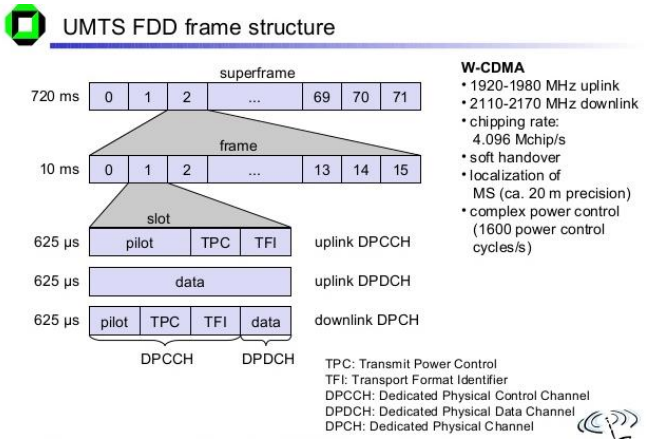


Figure. 4

A radio frame consists of 38400 chips and has a duration of 10ms. Each time slot consists of 2560 chips , which roughly equals 666.6µs.[1] The occupied bandwidth per W-CDMA channels is 4.4 to 5 MHz.

UTRAN

Figure. 5 shows the basic architecture of the UTRA network (UTRAN; 3GPP, 2002b). this consists of several radio network subsystems(RNS). Each RNS is controlled by **radio network controller (RNC)** and comprises several components that are called node B. An RNC in UMTS can be compared with the BSC; a node B is similar to a BTS. Each node B can control several antennas which make a radio cell. The mobile device , UE, can be connected to one or more antennas as will subsequently be explained in the context of handover. Each RNC is connected with the core network (CN) over the interface I_u and with a node B over the interface **I_{ub}**. A new interface , which has no counterpart in GSM , is the interface **I_{ur}** connecting two RNCs with each other. The use of this interface is explained together with the UMTS handover mechanisms.

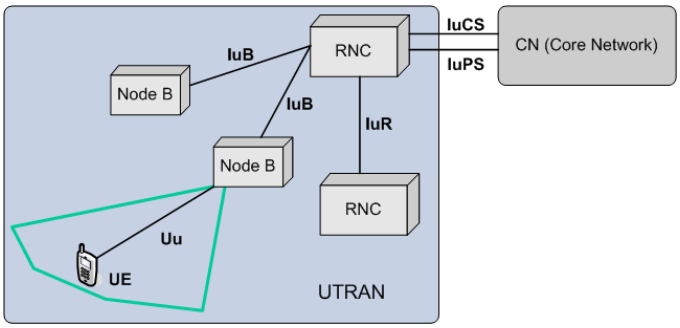


Figure. 5

Radio Network Controller

An RNC in UMTS has a broad spectrum of tasks as listed in the following:

- **Call admission control:** It is very important for CDMA system to keep the interference below certain level. The RNC calculates the traffic within each cell and decides, if additional transmissions are acceptable or not.

- **Congestion control:** During packet-oriented data transmission, several stations share the available radio resources. The RNC allocates band width to each station in cyclic fashion and must consider the QoS requirements.
- **Encryption/decryption:** The RNC encrypts all the data arriving from a fixed network before transmission over the wireless links and vice a versa.
- **ATM switching and multiplexing:** Typically, the connection between RNCs, nod Bs, and the CN are based on ATM. An RNC has two switch the connections to multiplex different data streams**.**
- **Radio resource controller:** The RNC controls all radio resources of the cells connected to it via nod B. This task includes interference and load measurements. The priorities of different connections have to be obeyed.
- **Radio barer set up and release:** An RNC has to set up, maintain, and release a logical data connection to a UE(the so called UMTS radio barer**.**
- **Code Allocation:** The CDMA codes used by a UE are selected by the RNC. These codes may vary during transmission.
- **Power control:** The RNC only performs a relatively loose power control. This means that the RNC influences transmission power based on interference values from other cells or even other RNCs.
- **Handover control and RNS relocation:** Depending on the signal strengths received by UEs and nod Bs, an RNC can decide if another cell would be better suited for certain connection. If a UE moves further out of the range of one RNC, a new RNC is responsible for the UE that has been chosen. This is called RNS relocation.
- **Management:** Finally, the network operator needs a lot of information regarding the current load, current traffic, error states etc. to manage its network. The RNC provides interfaces for this task, too.

Summary

3GPP UMTS, the Universal Mobile Telecommunications System is the third generation (3G) successor to the second generation GSM based cellular technologies which also include GPRS, and EDGE. Although UMTS uses a totally different air interface, the core network elements have been migrating towards the UMTS requirements with the introduction of GPRS and EDGE. In this way the transition from GSM to the 3G UMTS architecture did not require such a large instantaneous investment.

UMTS uses Wideband CDMA (WCDMA / W-CDMA) to carry the radio transmissions, and often the system is referred to by the name WCDMA. It is also gaining a third name.

Acknowledgment

This research was supported by Institute of Innovation in Technology and Management. We thank our colleagues of IITM who provided us insight and expertise that greatly assisted the research, although they may not agree with all of the interpretations of this paper. We thank Mr. Ajay Phogat, Assistant Professor at IITM, for assistance with his technical knowledge and guidance in formation of this paper.

References

[1] 3GPP (2000) General UMTS architecture , 3rd Generation Partenership Poject, 3G TS 23.101 3.3.0(2000-12).

[2] 3GPP (2002a) 3rs Generation Partenership Project , http://www.3gpp.org/.

[3] 3GPP (2002b) UTRAN overall description , 3rd Generation Partenership Project, 3GPP TS 25.401 V3.10.0 (2002-06)

[4] ETSI (1991a) Bearer services supported by a GSM PLMN, European Telecommunications Standards Institute, GSM recommendation 01.02.

[5] ETSI (1991b) General description of a GSM PLMN, European Telecommunication Standards Institute, GSM recommendation 01.02.

[6] ESTI (1991c) Subscriber Identity Modules , Functional Characteristics , European Telecommunications Standards Institute , GSM recommendation 02.17.

Applications of Data Mining in Higher Education

Sahil Gupta and Rahul Ahuja

***Abstract*- Data analysis plays important role in support for decision making, without considering the type of industry it is done for. There are n number of disciplines where data mining plays an important role. This paper proposes various means of data mining to improve the efficiency of the higher education system. Amongst the set of all the data mining techniques (clustering, decision tree, association) available, if any of the data mining technique is applied to the higher education system, it would help to improve the performance of the students and their life cycle management, selection of courses to measure their retention rate. This is an approach to examine the effect of using data mining techniques in higher education.**

Keywords: ERP, Decision support, Data Warehousing, OLAP, Data Mining, Applications in Higher Education.

1. Introduction

Higher education institutions today, are more concerned in predicting the appropriate paths for the students and the alumni. They are to decide as to which career a particular student is suitable to. Today, one of the biggest challenges that educational institutions face is the explosive growth of educational data and to use this data to improve the quality of managerial decisions. Data mining techniques are analytical tools that can be used to extract meaningful knowledge from these large data sets. To face these challenges different systems are used such as ERP, DWH etc.

1.1 Enterprise Resource Planning (ERP)

It is the integration of the internal as well as the external management information of entire organization. manufacturing, sales, service information and most important its customer relationship management like finance/accounting information, manufacturing. ERP systems automate all of these activities and control the flow of the information among all the departments of the organization. It is also responsible for the managing the out flow of the information. ERP systems have a common database. It also supports decision making and thus proves to be helpful to the businessmen while making decisions. ERP opened new horizons in almost all sectors. In this paper, we are mainly concerned about educational sector. Education is one of India's prime Industries today. Its size is more than domestic software industry or automobile industry. Due to exponential growth of educational sector, educational institutes have now become complex organizations. The institutes today are not only into delivering the education to the students but also they are now inclining more towards delivering other services to the students such as placements, managing internal operations like smooth conducting of classes or recruitment and motivation of human resources like faculty and staff, financial and cash flow planning, co-ordination with regulatory and statutory authorities. In addition to that, institutes are also subjected to the vagaries of market forces due to stiff competition and demanding customers (students and corporate). Educational institutes need modern management practices and state of the art technology to manage their internal and external operations. Realizing such demands, software industry started developing automated solutions for educational administration.

1.2 OLTP and OLAP ERP's

OLTP (On-Line Transaction Processing) [6] has been developed as complete ERP solution for academic institution especially targeted to engineering colleges. The main objective of the OLTP is to give the students the robust and the advanced technology atmosphere.it helps us in managing a number of students by the use of only one integrated system. I t works the same way as the modern knowledge portal websites which are being developed to help students in giving the online tests and also to get the desired results in a short span of time. The faculty at the same time could build their tests. and publish them. Student's

interactions with online learning environments enable them to access online exercise work, to know their mistakes and to get teacher's comments etc. This real-time infrastructure monitoring and data protection solution comes with various functionalities and modules like SMS integration, e-mail communication, biometric support, payment gateway integration, shared mail folders XML gateway integration, LDAP address book, e-mail group address and mobile mail access. These modules can be customized to the specific requirement of the institute.

Table 1: Differences between OLTP and OLAP

	OLTP	OLAP
Application	Operational: ERP, CRM, legacy apps,	Management information System, Decision Support System
Typical Users	Staff	Managers, Executive
Horizon	Weeks, Months	Years
Refresh	Immediate	Periodic
Data model	Entity Relationship	Multi- Dimensional
Schema	Normalized	Star
Emphasis	Update	Retrieval

1.3 Data Warehousing

Data Warehousing is an information delivery system. system. It is another solution specifically designed for query and analysis of information related to any educational institute. It is relational database approach rather than traditional transaction processing. Data warehouse is categorized based on its data storage. The main source of the data is cleaned, transformed, catalogued and made available for use by managers and other business professionals for data mining, online analytical processing, market research and decision support. However, the means to retrieve and analyze data, to extract, transform and load data, and to manage the data dictionary are also considered essential components of a data warehousing system. In an educational institute it plays a very important role. Its main benefits in an educational institute are listed as follows.

 It provides an integrated and total view of an Institute.

 It makes the institute's current and historical information easily available for the decision making.

 It provides the facility to students to get their different subject notes from a web enabled database.

 It provides the information about student's attendance.

 Students can get their results easily and very quickly.

 It helps to provides information about faculty like how many members are their in all the different departments etc. Overall we can say that data

Overall, we can say that data warehousing just simplifies a complex system to a simple and easily accessible system. Data Warehouse maintains its function in three layers staging, integration and access. Staging is used to store raw data for developers. The integration layer is used to integrate data and to have a level of abstraction from users. The access layer is for getting out for users.

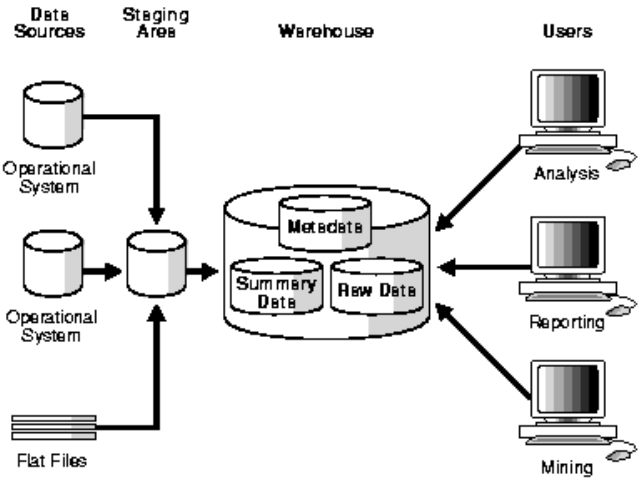


Fig. 1. Data Warehouse architecture

1.4 Data Mining

Data Mining is an interdisciplinary field of astronomy, business, computer science, economics and others to discover new patterns from large data sets. The actual data mining task is to analyze large quantities of data in order to extract previously unknown patterns such as groups of data records (cluster analysis), unusual records (anomaly detection) and dependencies (association rule mining)

These patterns can then be seen as a kind of summary of the input data, and used in further analysis. Data mining tasks can be classified as

 Anomaly detection (Outlier/change/deviation detection): The identification of unusual data records, that might be interesting or data errors which require further investigation.

 Association rule learning (Dependency modeling): Searches for relationships between variables. For example a supermarket might gather data on customer purchasing habits. Using association rule learning, the supermarket can determine which products are frequently bought together and use this information for marketing purposes.

 Clustering: It is a task of discovering groups and structures in the data that are in some way or another "similar", without using known structures in the data.

 Classification: It is the task of generalizing known structure to apply for new data. For example, an email program might attempt to classify an email as legitimate or spam.

 Regression: It attempts to find a function which models the data with the least error.

 Summarization: It providing a more compact representation of the data set, including visualization and report generation.

2. Data Mining Techniques

In this section, the main data mining techniques [9] used to analyze data are elaborated Fig. 2. Picture showing the partition of students in clusters There are many clustering algorithms reported in the literature. One has to take many decisions while choosing the appropriate algorithm for a particular problem. Some of them are Connectivity models: These are based on distance connectivity. Centroid models: This algorithm represents each cluster by a single mean vector.

Distribution models: Clusters are modeled using statistic distributions, such as multivariate normal distributions used by the Expectation-maximization algorithm.

Density models: It defines clusters as connected dense regions in the data space.

Subspace models: Clusters are modeled based upon both cluster members and relevant attributes.

2.2 Decision Tree

Decision tree is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm. Decision trees are commonly used in operations research, specifically in decision analysis, to help identify a strategy most likely to reach a goal. Another use of decision trees is as a descriptive means for calculating conditional probabilities. Decision trees can be used to analyze the admission criteria of an institute as shown in fig.3. Decision trees are simple to understand and interpret and moreover they give good results even with small data. This approach may not be suitable for data including categorical variables with different number of levels.

2.3 Factor Analysis

Factor analysis is a statistical method used to describe variability among observed, correlated variables in terms of a potentially lower number of unobserved, uncorrelated variables called factors.

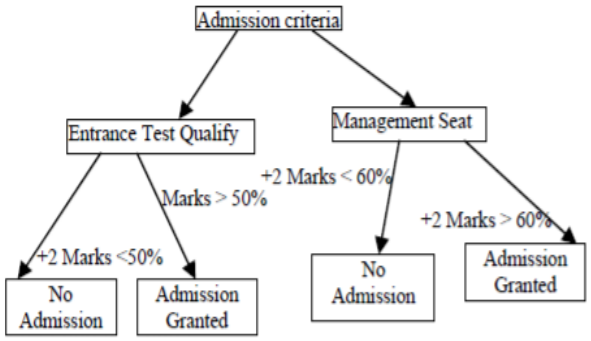


Fig. 3. Decision tree based upon admission criteria

In other words, it is possible, for e.g. variations in three or four observed variables mainly reflect the variations in fewer such unobserved variables. Factor analysis searches for such joint variations in response to unobserved latent variables. The observed variables are modeled as linear combinations of the potential factors, plus "error" terms. The information gained about the interdependencies between observed variables can be used later to reduce the set of variables in a dataset. Factor analysis originated in psychometrics and used in behavioral sciences, social sciences, marketing, product management, operations research, and other applied sciences that deal with large quantities of data. It can be of two

types, Exploratory factor analysis (EFA) and Confirmatory factor analysis(CFA)

Exploratory factor analysis (EFA): is used to uncover the underlying structure of a relatively large set of variables. The researcher's a priori assumption is that any indicator may be associated with any factor. This is the most common form of factor analysis.

Confirmatory factor analysis (CFA): seeks to determine if the number of factors and the loadings of measured (indicator) variables on them conform to what is expected on the basis of pre-established theory. Indicator variables are selected on the basis of prior theory and factor analysis is used to see if they load as predicted on the expected number of factors.

2.4 Regression Analysis:

In statistics, regression analysis includes techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables. More specifically, regression analysis helps us to understand how the typical value of the dependent variable changes when any one of the independent variables is varied, while the other independent variables are held fixed. Most commonly, regression analysis estimates the conditional expectation of the dependent variable with respect to independent variables. In such cases, the estimation target is a function of the independent variables called the regression function. In regression analysis, it is also of interest to characterize the variation of the dependent variable around the regression function, which can be described by a probability distribution. Regression analysis is widely used for prediction and forecasting, It is used to explore relationship between independent variables and dependent variable, Regression methods mainly used are linear regression and ordinary least squares regression.

3. Applications of Data Mining in Higher Education

There are many application areas of data mining like customer analytics, Agriculture, banking, Security Applications, Educational data mining, Mass surveillance, Privacy preserving etc. The main concerned area is about data mining applications in educational systems. Educational Data Mining (EDM) is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in.[A key area of EDM is mining student's performance. Another key area is mining enrollment data. Key uses of EDM include predicting student performance and studying learning in order to recommend improvements to current educational practice. EDM can be considered one of the learning sciences, as well as an area of data mining. The main applications of EDM are listed as follows

3.1 Analysis and Visualization of Data

It is used to highlight useful information and support decision making. In the educational environment, for example, it can help educators and course administrators to analyze the students' course activities and usage information to get a general view of a student's learning. Statistics and visualization information are the two main techniques that have been most widely used for this task. Statistics is a mathematical science concerning the collection, analysis, interpretation or explanation, and presentation of data. It is relatively easy to get basic descriptive statistics from statistical software, such as SPSS. Statistical analysis of educational data (logs files/databases) can tell us things such as where students enter and exit, the most popular pages students browse, number of downloads of e-learning resources, number of different pages browsed and total time for browsing different pages. It also provides knowledge about usage summaries and reports on weekly and monthly user trends, amount of material students might go through and the order in which students study topics, patterns of studying activity, timing and sequencing of events, and the content analysis of students notes and summaries. Statistical analysis is also very useful to obtain reports assessing how many minutes student worked, number of problems he resolved and his correct percentage along with our prediction about his score and performance level. Visualization uses graphic techniques to help people to understand and analyze data. There are several studies oriented toward visualizing different educational data such as patterns of annual, seasonal, daily and hourly

user behavior on online forums. Some of such investigations are statistical graphs to analyze assignments complement, questions admitted, exam score, student tracking data to analyze student's attendance, results on assignments and quizzes, weekly information regarding students and group's activities.

3.2 Predicting Student Performance

In this case, we estimate the unknown value of a variable that describes the student. In education, the values normally predicted are student's performance, their knowledge, score, or marks. This value can be numerical/continuous (regression task) or categorical/discrete (classification task). Regression analysis is used to find relation between a dependent variable and one or more independent variables. Classification is used to group individual items based upon quantitative characteristics inherent in the items or on training set of previously labeled items. Prediction of a student's performance is the most popular applications of DM in education. Different techniques and models are applied like neural networks, Bayesian networks, rulebased systems, regression, and correlation analysis to analyze educational data. This analysis helps us to predict student's performance i.e. to predict about his success in a course and to predict about his final grade based on features extracted from logged data. Different types of rule-based systems have been applied to predict student's performance (mark prediction) in an elearning environment (using fuzzy-association rules). Several regression techniques are used to predict student's marks like linear regression for predicting student's academic performance, stepwise linear regression for predicting time to be spent on a learning page, multiple linear regression for identifying variables that could predict success in colleges courses and for predicting exam results in distance education courses.

3.3 Outlier Analysis:

According to Grubbs [11] Outlier can be defined as "An outlying observation, or outlier, is one that appears to deviate markedly from other members of the sample in which it occurs". Outlier detection has been used to detect and, where appropriate, remove anomalous observations from data. Outlier detection can identify system faults and fraud before they escalate with potentially catastrophic consequences. There are three fundamental approaches for outlier detection.

- Type 1 - Determine the outliers with no prior knowledge of the data. This is essentially a learning approach analogous to unsupervised clustering. The approach processes the data as a static distribution, pinpoints the most remote points, and flags them as potential outliers.
- Type 2 - Model both normality and abnormality. This approach is analogous to supervised classification and requires pre-labeled data, tagged as normal or abnormal.
- Type 3 - Model only normality (or in a few cases model abnormality). This is analogous to a semisupervised recognition or detection task. It may be considered semi-supervised as the normal class is taught but the algorithm learns to recognize abnormality.

3.4 Grouping Students

In this case groups of students [10] are created according to their customized features, personal characteristics, etc. These clusters/groups of students can be used by the instructor/developer to build a personalized learning system which can promote effective group learning. The DM techniques used in this task are classification and clustering. Different clustering algorithms that are used to group students are hierarchical agglomerative clustering, K-means and model-based clustering.

3.5 Planning and Scheduling

Planning and scheduling is used to enhance the traditional educational process by planning future courses, course scheduling, planning resource allocation which helps in the admission and counseling processes, developing curriculum, etc. Different DM techniques used for this task are classification, categorization, estimation, and visualization. The main objective of using above techniques is academic planning, predicting alumni pledges and creating meaningful learning outcome typologies. Decision trees, link analysis and decision forests have been used in course planning to analyze enrollee's course preferences and course completion rates in extension education courses. Classification, prediction, association-rule analysis, clustering, etc have been compared to discover new explicit knowledge that could be useful in the

decision-making process in higher learning institutions. Educational training courses have been planned through the use of cluster analysis, decision trees, and back-propagation neural networks in order to find the correlation between the course classifications of educational training. Decision trees and Bayesian models have been proposed to help management institutes to explore the probable effects of changes in recruitments, admissions and courses.

3.6 Enrollment Management

This term is frequently used in higher education to describe well-planned strategies and tactics to shape the enrollment of an institution [12] and meet established goals. Enrollment management is an organizationalenable educational institutions to exert more influence over their student enrollments. Such practices often include marketing, admission policies, retention programs, and financial aid awarding. Strategies and tactics are informed by collection, analysis, and use of data to project successful outcomes. Activities that produce measurable improvements in yields are continued and/or expanded, while those activities that do not are discontinued or restructured. Competitive efforts to recruit students are a common emphasis of enrollment managers. The numbers of universities and colleges instituting offices of "enrollment management" have increased in recent years. These offices serve to provide direction and coordination of efforts of multiple offices such as admissions, financial aid, registration, and other student services. Often these offices are part of an enrollment management division.

Some of the typical aims of enrollment management include

 Improving yields at inquiry, application, and enrollment stages.

 Increasing net revenue, usually by improving the proportion of entering students capable of paying most or all of unsubsidized tuition.

 Increasing demographic diversity

 Improving retention rates

 Increasing applicant pools

4. Conclusion and Future Scope

In this paper, some important issues related to business community and education system are discussed along with their solutions. Data analysis plays an important role for any type of decision support irrespective of type of industry. Data warehousing and data mining methods for data analysis are explained in detail. Main core of this paper is to review role of data mining techniques in education system. Educational Data Mining has been introduced as an upcoming research area, Thus number of specific tools specially developed for applying DM algorithms in educational data/environments are emerging day by day.DM techniques in educational organizations help us to learn student performance, student behavior, carefully designing course curriculum, to motivate students and to group student depending upon various parameters.

It is observed that current DM tools are too complex for educators to use and their features go well beyond the scope of what an educator may want to do. One possible solution is the development of tools that use a default algorithm for each task and parameter-free DM algorithms to simplify the configuration and execution for non-expert users. Secondly, the DM tool has to be integrated into the e-learning environment so that results obtained with DM techniques could be easily and directly applied. Moreover, current tools for mining data pertaining to a specific course/framework may be useful to their developers only. There are no general tools or reusing tools that can be applied to any educational system. Therefore, a standardization of input data and output model are needed. Data mining techniques are useful in student marketing, selection revenue analysis, predicting student performance, planning of courses and result analysis. So, it has a wide array of applications for the higher education sector.

References

- www.dynelytics.com
- www.ijcsi.com
- Research paper by
  Monika Goyal1 and Rajan Vohra2
  1, 2 CSE Department, BahraUniversity,
  Waknaghat, H.P 173234, India
- www.semanticacholar.com.

SMART CITY SMART VILLAGE

1. INTRODUCTION

A smart city is a urban development vision to integrate multiple information and communication technology and Internet of Things(IOT) solutions to a secure fashion to manage a city's assests.The rapid development of information technology(IT) has brought forward a hyper connected society in which objects are connected to mobile devices and the internet and communicate with one another .In the 21 st century ,we want to be connected with anything anytime and anywhere, which is already happening in various places around the world . Recently, many local governments have been aiming to implement an IoT-based smart city through the construction of a test bed for IoT verification and an integrated infrastructure. This movement also corresponds to the creative economy that is emphasized by the Korean government. A Smart Village is a bundle of dozens of services delivered effectively to the residents and businesses in an efficient manner. These services could be location specific depending on the demography of the village and occupations of the residents. These services such as Power, Water, Buildings, Retail, Health care, etc. were built several decades ago. New designs, technologies and management models should be used to upgrade the existing ones and in building the new ones.

This requires standardization, use of IT and sensor networks. Requires strategy, integrated planning and above all monitoring and execution of the activities using appropriate governance models.

The concept of Smart cities

.1 Smart Buildings

The homes and buildings can be made smart by the use of sensors and cameras. These will produce real-time data which can be analyzed to take necessary actions. For example, sensors installed in a home can detect smoke and hence start the water sprinklers automatically to combat the fire. Similarly, the sensors can monitor the usage of electricity in the home or building and switch the lights off when not in use. The security of the building can be monitored using cameras and appropriate alerts can be generated in case of any anomalies. The water levels and pressure can be measured in the water tanks and pipes and used to refill the tanks when necessary as well as detect any faults in the pipes.

2. Smart Weather and Irrigation

Accurate weather information can be of great use to the people of the village. As we know, the majority of population in villages engages in agriculture for their living. The use of environmental sensors to predict weather forecasts can help the farmers to a large extent. Many farming activities like sowing, irrigation and

harvesting depend on the weather. Smart irrigation systems can make use of sensors in the fields and remote satellite data to ensure the optimal use of available water resources. If it is going to rain the next day, then watering the fields on that day makes no sense. The level of water in the dams and canals can also be monitored using sensors and
it can be used to predict the future need of water.

3.**Smart Farming**

As Agriculture is the backbone of all villages, the farmers need to benefit the most from the system of IoT and Smart
villages. There needs to be the tracking of the farm produce from the farm to the table. The whole chain of activities can be monitored and improved using data from sensors and other sources. The people involved in the process are the growers, processors and packers, storage and transport service providers, distributors, wholesalers and retailers.

4. Smart Dairy

The secondary occupation of a large number of farmers is number of farmers is rearing cattle for dairy products .The use sensors and camera is to ban or shelter can help the farmers in better management of their work .Any changes can be reported instantly through alert messages and

required measures can be taken favorable temperature for the cattle can be maintained using smart devices . The food water and health necessities of the cattle can also be monitored in a similar fashion. Grazing the cattle in the open fields is a risky thing if there is no one to supervise it. The use of sensors in the fields can eliminate the job of supervision by a human and it can be done remotely by the farmers.

5. Smart Healthcare

Smart health services are needed to improve the quality of life in the villages. The village dispensaries and hospitals need advanced devices which are connected to each other and the doctors. The beds in hospital can be embedded with sensors which can detect various changes in the patient including its movements, heartbeat, blood flow from the wounds and body temperature etc. These reports along with the data generated by various machines like X -rays, CT scans etc. can be sent to the doctor directly. Such services will upgrade the health care sector of the villages.

6. Smart Education

Education is the basic means to implement all the advancements in life. Educating people about the use of new technologies facilitates better implementation. It can be the force behind reducing the digital

-divide which is far more prevalent in villages than the cities. The whole idea of Smart villages revolves around its people and how efficiently they make use of the components of a Smart village. They can be educated to participate in each and every activity of the village leading to a better lifestyle for its people. Dealing with children and teenagers becomes easier when we educate them in an interesting way. Video games and lectures fascinate most of the children and can help them learn in an interactive manner rather than reading the text
-books in the classrooms. Internet of Things (IoT) brings together different technologies like Internet, Mobile and smart devices and hence assists in the learning process. The use of LCD screens and interactive videos can foster the learning in children and even adults. These can be used to educate them to use the facilities provided in the Smart villages in the best way. The village schools can be equipped with Internet and other devices and learning can be made a fun activity turning the schools into Smart schools.

CONCLUSION

A few years ago, the idea of Internet of Things and Smart cities used to be considered as a future possibility. But it has become a reality today, thanks to the technological advancements. Many countries have deployed the job of

turning their cities into Smart cities to many organizations. The optimal use of available resources is the need of the hour. Ever-increasing population has restrained the resources and their usage. IoT combines the benefits of multiple technologies to realize the idea of intelligent devices in a city. This idea can be extended to the villages as well, improving the quality of life of the residents. As the villages have slightly differences and aims to provide solutions for the same. Various areas of interest have been explored and suggestions are also provided.

Cyber Security

Ajay Phogat
Department of IT
Institute of Innovation in Technology &
Management

Rohit kumar,
Student-BCA
Institute of Innovation in Technology &
Management

INTRODUCTION

Our world is increasingly connected through sophisticated networks, internet portals for commerce, mobile devices, tablets, and other innovative tools providing opportunities for economic growth, innovation and convenience. As businesses, governments and individuals become more reliant on these connections valued assets are increasingly accessible and cyber security threats multiply.

Cyber criminals are more sophisticated, targeted and better funded than ever. And crime follows monetization opportunities. There is an emerging correlation between the size of an organization and the type of data targeted. Credit card payments and authentication credentials tend to be typical targets within smaller organizations. Data of strategic significance such as trade secrets and other intellectual property are growing targets within larger organizations.

On the other hand the cost or risk of engaging in cyber crime is often very low relative to the pay-off. Attribution and "chain of custody" issues make prosecution by law enforcement difficult. In some cases even when criminals are prosecuted successfully the penalties are not significant enough to be a deterrent.

The Challenge

No matter what strategy is adopted breaches will occur. It is nearly impossible to take advantage of our connected without being at risk. Defensive technologies such as firewalls, passwords, encryption, physical barriers, and authentication mechanisms are important to maintain but alone have not been effective in eliminating breaches or predicting where the next attack will occur. Their value as stand-alone security measures will be of limited use in fighting increasingly sophisticated, innovative and well-funded cyber criminals.

The emerging challenge is to find more predictive methods of identifying threats mitigating their impact and managing an agile cyber security operation that will both creatively and effectively maintain protection.

important to recognize that:

- It is not economical to protect every piece of data and every asset to the same extent.
- A balance between the right to privacy with the need to protect nations, enterprises and individuals from intrusions must be negotiated.

The challenge is great and requires fresh ways to blend people, processes, technology and shared data to protect societies from emerging threats to security.

Designing a Resilient Enterprise

For any individual or organization to thrive over a sustained period some level of resilience is required. How does one build resilience in a rapidly changing environment where emerging threats are taking on increasing sophistication and severity?

In this discussion an enterprise is defined as a unit of organization or activity. So a company, business, government entity or not-for-profit organization may be an enterprise. Every enterprise has a mission and a need to maximize results towards that mission whether it is social, economic, diplomatic or otherwise. Strategic activities within the enterprise align with that mission to facilitate its success. Cyber security is one of those strategic activities to be managed holistically as its effectiveness broadly impacts the enterprise's ability to carry out its mission. It impacts how we interact with customers design new products, market services, manage operations and set policies.

As a strategic decision cyber security becomes the charge of the enterprise as a whole and is considered through a risk versus investment lens rather than simply a as technology purchase. Organizations don't typically have the resources to protect every asset and some assets do not warrant as much protection as others. At the same time, organizations cannot afford to take an approach to security.

Most organizations collect data internally representing one data source. Increasingly, organizations are combining their selected data with that of other trusted public and private sources discovering that the predictive value of broader based data analytics increases exponentially. As the Titan example demonstrates analysis of larger data sets reveals correlations and patterns of current threats that a single source simply cannot. Additionally, it allows emerging threat vectors and command and control mechanisms to be quickly identified so that each participating organizations may adjust security measures to mitigate these threats and protect precious assets.

Maintaining Enterprise Resilience

Once priorities are set and investment decisions are made, the cyber security operations structure must be able to effectively implement and administer protection plans. Agility and flexibility are hallmarks of an effective cyber security operation, meeting daily demands while addressing and emerging threats.

Policies, compliance standards, workflows and established processes guide daily operations. But real-time actionable data will drive security operations decisions in a resilient enterprise. Advanced data visualization techniques previously mentioned allow administrators to monitor daily activities while recognizing the nuances of abnormal behaviors.

Speed of detection and response are critical when trying to limit the damage caused by a breach. When a problem is detected defined operational workflows clear roles and responsibilities, policies, decision-making authority and adequate training guide an administrator's response. Research indicates that organizations with a well-defined incident response plans are better able to respond effectively to a breach. Plans outline procedures for minimizing damage or loss collecting data on the incident, preserving evidence, mitigating the vulnerability on a temporary or permanent basis and communicating the incident within the organization. Both NIST and ENISA provide guidance on creating effective incident response plans.

Cyber security is the responsibility of the enterprise rather than a single team. As such, building a culture that supports security standards compliance, teaches its members how to recognize abnormal behavior. Furthermore organizations must reward participation in security programs. People are the eyes and ears of the daily operations providing broad situational awareness and proactive protection at all levels of the enterprise.

Policy Implications and Incentives

Creating policy to mitigate cyber threats while preserving privacy and limiting government intervention to a comfortable level is a tricky balancing act. But there are opportunities to influence future preparedness through forward thinking policy development.

Investment in Innovation will be a critical step to maintaining security and competitiveness on a global scale while limiting damage and other cyber criminal activity. The following areas are important targets for investment as their correlation with threat prediction and damage control makes these especially valuable opportunities.

- Real-time threat detection and data analysis tools – many tools exist today but their level of sophistication and wide spread adoption must continue to grow to provide more comprehensive protection.

- Big Data – to effectively compile and correlate large volumes of data, new technologies and algorithms will be required.

- Visualization tools – related to big data opportunities are visualization techniques: creative visual presentations of data that quickly differentiate warning signs from normal operating behaviors.
- Emerging technologies that contribute to resilience more robust protection and attribution of cyber crimes.

Data Sharing: As real-time data analysis for decision-making is a pillar for future cyber security strategies breaking down barriers for security data sharing amongst trusted partners is a necessary next step in predicting and mitigating emerging threats. Policies may provide incentives for participation define disclosure boundaries and rules of engagement between enterprises and nations and encourage and create networks of trusted partners.

Law Enforcement: Laws governing cyber crime and resources to prosecute criminals are inadequate to address the attacker's sophistication and the damage caused. Many agencies such as the FBI and Europol appropriately focus priorities on child protection, terrorism and counter-intelligence with limited budgets to achieve their missions. However, the loss of industry's intellectual property and trade secrets will also have a lasting and severe economic impact on these nations.

Developing an Educated Cyber Workforce: The need for skilled professionals and technicians to address cyber security continues to grow. However demand for these individuals exceeds the supply and the problem is projected to grow in the future. Policies may provide incentives for students to select cyber education paths and create broader awareness for the opportunities that exist in this growing industry. Retraining workers from declining industries may also represent an opportunity to meet future demands.

Supporting Cyber Hygiene: Creating access to security tools and best practices will be important to fighting cyber crime for individuals, enterprises, nations and the world. Connectedness requires that security solutions be broadly implemented to be effective. Policies may influence access to these tools provide education on their use create incentives for use and compliance with standards and create a culture of responsibility for security.

Privacy, Reporting and Government Role: Several broader policy issues that govern our collective approach to cyber security have large implications for the future:

- The right to privacy by the individual and the enterprise – when should collective security interests and protection be more important than individual privacy rights?
- The roles government should play in cyber security.
- Reporting requirements for security breaches.
- Lack of consistency in laws and requirements between nations and severity of penalties.

These are complex and sometimes controversial policy issues but incentives established by new policies may have far reaching influence on the level of protection and the approaches we can take to protecting individuals, enterprises and nations from cyber crime of the future.

References

1. Interpol. (2013). Cybercrime. Retrieved from http://www.interpol.int/Crimeareas/Cybercrime/Cybercrime/

2. 2012 Data Breach Investigations Report conducted by the Verizon RISK Team (2012). Retrieved from http://www.verizonbusiness.com/about/events/2012dbir/

3. Merriam-Webster Dictionary. (2013). Retrieved from http://www.merriamwebster.com/

4. Changing the Game: Key findings from the PWC Global State of Information Security Survey 2013 (2013). Retrieved from http://www.pwc.com/

5. Finding a Strategic Voice: Insights from the 2012 IBM Chief Information Security Officer Assessment (2012). Retrieved from http://www.ibm.com/

6. Bilby, E. (2012, December 17). EU could make firms disclose network security breaches. Reuters. Retrieved from http://reuters.com/

7. U.S. Department of Commerce, National Institute of Standards and Technology (2012). Computer Security Incident Handling Guide – Recommendations of the National Institute of Standards and Technology (Special Publication 800-61, Revision 2). Retrieved from http://www.nist.gov/index.html

8. European Network and Information Security Agency. (2006, May). A Step-ByStep Approach On How To Set Up A CSIRT. Retrieved from http://www.enisa.europa.eu/

9. European Network and Information Security Agency. (2012, December). Consumerization of IT: Risk Mitigation Strategies – Responding to the Emerging Threat Environment. Retrieved from http://www.enisa.europa.eu/

Abstarct

Software development life cycle is a process used by software company to design, develop and test softwares. It aims to produce high quality software to meet customer specification.testing plays important role in making software.testing is the process of evaluating the system or its component with the intend to find whether it satisfies the specific requirements or not.in my paper I have explained various phases,different types of testing and importance of testing in software development life cycle.

Introduction

Software testing plays important role in software development life cycle. But before discussing importance of testing, I would like to discuss different phase of software development cycle.

Phase 1 Requirement gathering and analysis : Requirements are gathered in this phase.In this phase the main focus is on project managers and stake holder and meetings are done in order to determine the requirements like: Who is going to use the system? What data should be input into system?. Finally, a software requirement specification document is created.

Phase 2 Design : In this phase the system and software design is prepared from the requirement specifications which were studied in the first phase. System Design helps in specifying hardware and system requirements and also helps in defining overall system architecture.

Phase 3 Implementation / Coding : This is the important and longest phase in the software life cycle. This part is the main focus for the developer.

Phase 4 Testing : After the code is developed it is tested against the requirements to check that the product is actually meet specific requirements or not. In this phase , all types of functional and non-functional testing are also done.

Phase 5 Deployment : After successful testing the product is delivered / deployed to the customer for their use.

Phase 6 Maintenance : Once when the customers starts using the developed system then the actual problems comes up and needs to be solved from time to time. This process where the care is taken for the developed product is known as maintenance.

Importance of Testing

Testing is important because software bugs could be expensive or even dangerous. Software bugs can potentially cause monetary and human loss, history is full of such examples.

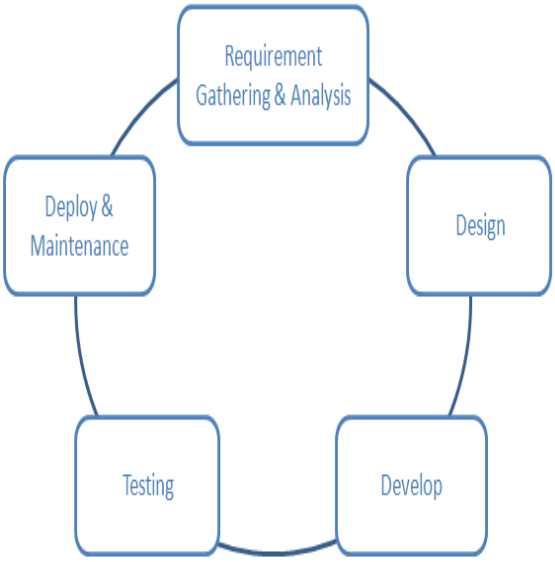
- In April 2015, Bloomberg terminal in London crashed due to software glitch affected more than 300,000 traders on financial markets. It forced the government to postpone a 3bn pound debt sale.
- Nissan cars have to recall over 1 million cars from the market due to software failure in the airbag sensory detectors. There has been reported two accident due to this software failure.
- Starbucks was forced to close about 60 percent of stores in the U.S and Canada due to software failure in its POS system. At one point store served coffee for free as they unable to process the transaction.
- Some of the Amazon’s third party retailers saw their product price is reduced to 1p due to a software glitch. They were left with heavy losses.

Various types of testing are:-

1. **Age Testing:** Type of testing which evaluates a system's ability to

perform in the future. The evaluation process is conducted by testing teams.

2. **Basis Path Testing:** A testing mechanism which derives a logical complexity measure of a procedural design and use this as a guide for defining a basic set of execution paths. It is used by testing teams when defining test cases.
3. **Configuration Testing:** Testing technique which determines minimal and optimal configuration of hardware and software, and the effect of adding or modifying resources such as memory, disk drives and CPU.
4. **Dependency Testing:** Testing type which examines an application's requirements for pre-existing software, initial states and configuration in order to maintain proper functionality. It is usually performed by testing teams.
5. **Error-Handling Testing:** Software testing type which determines the ability of the system to properly process erroneous transactions. It is usually performed by the testing teams.
6. **Functional Testing:** Type of black box testing that bases its test cases on the specifications of the software component under test. It is performed by testing teams.
7. **Gorilla Testing:** Software testing technique which focuses on heavily testing of one particular module. It is performed by quality assurance teams, usually when running full testing.
8. **Load Testing:** Testing technique that puts demand on a system or device and measures its response. It is usually conducted by the performance engineers.
9. **Mutation Testing:** Method of software testing which involves modifying programs' source code or byte code in small ways in order to test sections of the code that are seldom or never accessed during normal tests execution. It is normally conducted by testers.
10. **Regression Testing:** Type of software testing that seeks to uncover software errors after changes to the program (e.g. bug fixes or new functionality) have been made, by retesting the program. It is performed by the testing teams.



Different phase of SDLC

11. **Smoke Testing:** Testing technique which examines all the basic components of a software system to ensure

that they work properly. Typically, smoke testing is conducted by the testing team, immediately after a software build is made .

12. **Stress Testing:** Testing technique which evaluates a system or component at or beyond the limits of its specified requirements. It is usually conducted by the performance engineer

References :

Different phase of SDLC available at: http://istqbexamcertification.com/what-are-the-software-development-life-cycle-sdlc-phases/

Importance of testing available at: http://www.guru99.com/software-testing-introduction-importance.html

Types of testing available at: http://www.guru99.com/types-of-software-testing.html

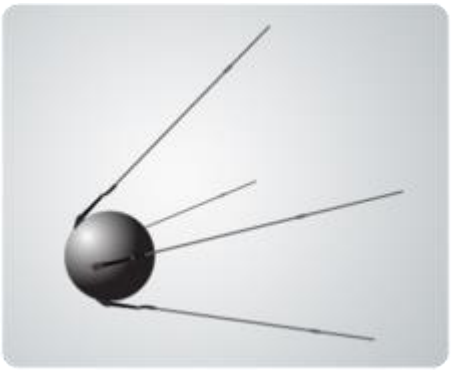
Conclusion :

Software development life-cycle is a structure imposed on the development of a software product. In my paper, I have discussed various phase of software development life cycle , role of testing ,importance of testing and different types of testing.

Tracking devices was introduced due to GPS. A GPS is a device by the help of which we can track and locate people , restaurants, various other things, etc. (GPS) **Global Positioning System** (**GPS**) is a space-based radionavigation system owned by the United States Government (USG) and operated by the United States Air Force (USAF)." It is a global navigation satellite system (GNSS) that provides geolocation and time information to a GPS receiver in all weather conditions, anywhere on or near the Earth where there is an unobstructed line of sight to four or more GPS satellites. Now this GPS is used as a tracking device to track and locate mobile users in different parts of the world. But , first we should know how did it originated.

HISTORY OF TRACKING DEVICES(GPS)

Originally designed for military and intelligence applications at the height of the Cold War in the 1960s, with inspiration coming from the launch of the Soviet spacecraft Sputnik in 1957, the global positioning system (GPS) - is a network of satellites that orbit the earth at fixed points above the planet and beam down signals to anyone on earth with a GPS receiver. These signals carry a time code and geographical data point that allows the user to pinpoint their exact position, speed and time anywhere on the planet. Transit was the first satellite system launched by the USA and tested by the There are



three parts to a GPS system: a constellation of between 24 and 32 solar-powered satellites orbiting the earth in orbits at an altitude of approximately 20000 kilometers, a master control station and four control and monitoring stations (on Hawaii, Ascension Islands, Diego Garcia and Kawajale) and GPS receivers such as the one in a car.

Each of the satellites is in an orbit that allows a receiver to detect at least four of the operational US Navy in 1960. Just five satellites orbiting the earth allowed ships to fix their position on the seas once every hour. In 1967 Transit was succeeded by the Timation satellite, which demonstrated that highly accurate atomic clocks could be operated in space. GPS developed quickly for military purposes thereafter with a total of 11 "Block" satellites being launched between 1978 and 1985.

However, it wasn't until the USSR shot down a Korean passenger jet - flight 007 - in 1983 that the Reagan Administration in the US had the incentive to open up GPS for civilian applications so that aircraft, shipping, and transport the world over could fix their positions and avoid straying into restricted foreign territory.

Upgrading the GPS was delayed by NASA space shuttle SS Challenger disaster in 1986 and it was not until 1989 that the first Block II satellites were launched. By the summer of 1993, the US launched their 24th Navstar satellite into orbit, which completed the modern GPS constellation of satellites - a network of 24 - familiar now as the Global Positioning System, or GPS. 21 of the constellation of satellites were active at any one time; the other 3 satellites were spares; in 1995 it was declared fully operational. Today's GPS network has around 30 active satellites in the GPS constellation.

Today, GPS is used for dozens of navigation applications, route finding for drivers, map-making, earthquake research, climate studies, and an outdoor treasure-hunting game known as geocaching.

HOW DOES GPS WORK?

satellites. The satellites send out microwave signals to a receiver where the built-in computer uses these signals to work out your precise distance from each of the four satellites and then triangulates your exact position on the planet to the nearest few meters based on these distances.

In fact, signals from just three satellites are needed to carry out this trilateration process; the calculation of your position on earth based on your distance from three satellites. The signal from the fourth

satellite is redundant and is used to confirm the results of the initial calculation. If the position calculated from distances to satellites “A-B-C” do not match the calculation based on “A-B-D” then other combinations are tested until a consistent result is obtained.

The process of measuring the distance from satellite to GPS receiver is based on timed signals. For example, at 16h45m precisely, the satellite may begin broadcasting its signal. The GPS receiver will also begin running the same random sequence at 16h45m local time, but does not broadcast the sequence. When the receiver picks up the signal from the different satellites, there will be a time lag, because the microwaves take a fraction of a second to travel from the satellite to the receiver. The time lag is easily converted into distance to each satellite. The slight difference between signals from each satellite is then used to calculate the receiver's position.

HOW IT WORKS ON MOBILE DEVICES:

Using GPS on your cell phone depends on the phone. Some phones use GPS only for 911 emergency calls. Sophisticated phones are able to receive and display location maps to show roads and more, and to give directions. These turn-by-turn programs are mostly JAVA-based and work with your cell provider's database.

COMMON USES FOR GPS IN PHONE:

- **Location Tracking**: Some employers use GPS-enabled phones to track their employees' locations, and some business offer location tracking services for GPS-enabled phones. The Wherifone locator phone provides GPS coordinates and can dial emergency phone numbers. Parents and caregivers can track the phone's location by phone or online and can receive notification if it leaves a designated "safe area." Wearable Environmental Information Networks of Japan has also introduced the Dog @ Watch, a GPS watch phone for children.
- **Turn-by-Turn Directions**: GPS-enabled phones with view screens can often display turn-by-turn directions as well as announce them through the phone's speaker. In general, companies that offer these services charge a monthly fee and use a database of maps to provide the directions. The services are only as good as their database -- outdated maps can provide inaccurate directions. Some turn-by-turn direction services include: TeleNav ViaMoto MapQuest Find Mesmart2Go, which requires a separate Bluetooth GPS receiver and a memory card Destinator SP, which is a software package for smartphones.
- **Outdoor Location Services**: Trimble outdoors offers maps and location-based services for hiking, mountain biking, geocaching and other outdoor activities.
- **Other Location-Based Services**: Some companies hope to deliver news, coupons, advertisements and other information to cell phone users based on their location.

GPS DEVICES:

GPS devices are used in various products some of them are as under:-

- Personal Tracker
- Car Gps
- Gps Devices
- Gps Tracker
- Gps Tracking System
- Portable Gps Device
- Gps Tracking Device
- Portable Gps System
- Car Navigation System
- Car Gps System
- Gps Modules
- Gps Navigation System
- Gps Receiver
- Active Gps Antenna
- Aviation Navigation Lights
- Portable Gps Tracker
- Waterproof Gps
- Gps Software

- Bluetooth Gps Receiver
- Gps Antenna
- Portable Gps Receiver
- Handheld Gps
- Gps Vehicle Tracking System
- Gnss Receiver

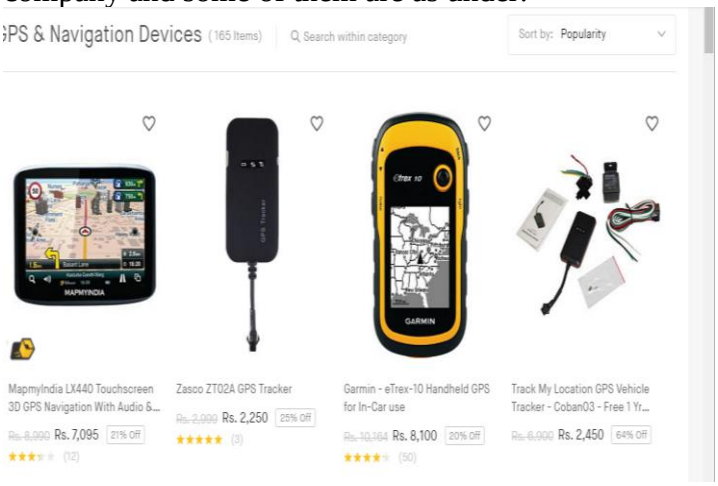
COMPANIES DEALING WITH GPS DEVICES:

There are various foreign companies too supporting GPS trading but some indian companies are also supporting its trade which are as under:-

- C.S. Aerotherm Pvt. Ltd.
- Global Tele Communication
- Mangal Security Products
- Atlas Comnet Pvt. Ltd.
- Link International
- Wift Technologies
- Applied Electro Magnetics Pvt. Ltd.
- Luxa Control Systems Pvt. Ltd.
- Sathya
- G. B. G. Engineering Tools
- Dvl Groups Ltd.
- Unitrack
- Opalsys It Solutions Pvt. Ltd

PRICE OF GPS DEVICES:

The price of GPS devices varies according to company and some of them are as under:-



And various sites such as Amazon , snapdeal , flipkart promote selling purchasing of GPS devices.

TRACKING DEVICES USED IN OTHER AREAS:

Tracking devices are used now a days in other areas, previously it was just to track shipment , in military to search enemy location except these now it is used commercially such as in booking taxi , food and reaching places if you don't know the road to reach a place.

Some example of such services are:-

- OLA Cabs , UBER, Meru Cabs, etc for cab services door to door or plce to place.
- Foodpanda, SWIGGY, Fasos, etc for booking food for lunch , dinner , snacks, breakfasts, etc.
- Hospital using GPS to reach to people / victims for providing support to reach hospital through ambulance.
- Google Maps to reach to places through small routes through car , metro , on foot, etc.
- Social apps such as Facebook, Snapchat , Instagram to tag location to share memories in social media.
- Use in games such as Pokemon GO, Minigami, etc. to provide open world gaming experience.

Tracking And Locating Mobile Users

Sarthak Dutta
sarthakd1995@gmail.com
IINTM, Janakpuri

Ankit Tayal
tayal1962@gmail.com
IINTM, Janakpuri

Rahul Kathait
RahulKathait123@gmail.com
IINTM, Janakpuri

ABSTRACT

Now a days, mobile tracking has become a part of life in this world of technology.
Without tracking we can't even place an order, book a taxi, locate nearby restaurants at our current location and reach to a desired place at time by the help of maps.
We use tracking for so many reasons but don't know from where it is originated. There's a phase called "Necessity is the mother of invention "and the necessity to cope up with technology getting advanced up day by day, tracking system was found which lead to make our daily tasks of life simple and easy without any problem.
Tracking system was developed due to invention of a device called as GPS. GPS stands for **Global Positioning System**. A GPS provides geo-location and time information to a GPS receiver in all weather conditions, anywhere on or near the Earth where there is an unobstructed line of sight to four or more GPS satellites. It uses satellites to transfer information related to a person's particular location, to forecast weather and to provide an image of the person's location around him. Tracking devices are also used in fitness devices such as FitBit watches to check heartbeat, distance covered, calories burned. Tracking devices are boon also to tech industry by providing ease to people and also by opening various business ventures operations some are google maps, booking applications like makemytrip, etc.

KEYWORDS: - **Global Positioning System, cyber cells, drones, e waste and 'no secrecy'.**

INTRODUCTION

Tracking devices was introduced due to GPS. A GPS is a device by the help of which we can track and locate people , restaurants, various other things, etc. (GPS) **Global Positioning System** (**GPS**) is a space-based radionavigation system owned by the United States Government (USG) and operated by the United States Air Force (USAF)." It is a global navigation satellite system (GNSS) that provides geolocation and time information to a GPS receiver in all weather conditions, anywhere on or near the Earth where there is an unobstructed line of sight to four or more GPS satellites. Now this GPS is used as a tracking device to track and locate mobile users in different parts of the world. But , first we should know how did it originated.

HISTORY OF TRACKING DEVICES(GPS)

Originally designed for military and intelligence applications at the height of the Cold War in the 1960s, with inspiration coming from the launch of the Soviet spacecraft Sputnik in 1957, the global positioning system (GPS) - is a network of satellites that orbit the earth at fixed points above the planet and beam down signals to anyone on earth with a GPS receiver.
These signals carry a time code and geographical data point that allows the user to pinpoint their exact position, speed and time anywhere on the planet.
Transit was the first satellite system launched by the USA and tested by the US Navy in 1960. Just five satellites orbiting the earth allowed ships to fix their position on the seas once every hour. In 1967 Transit

was succeeded by the Timation satellite, which demonstrated that highly accurate atomic clocks could be operated in space. GPS developed quickly for military purposes thereafter with a total of 11 "Block" satellites being launched between 1978 and 1985.

However, it wasn't until the USSR shot down a Korean passenger jet - flight 007 - in 1983 that the Reagan Administration in the US had the incentive to open up GPS for civilian applications so that aircraft, shipping, and transport the world over could fix their positions and avoid straying into restricted foreign territory.
Upgrading the GPS was delayed by NASA space shuttle SS Challenger disaster in 1986 and it was not until 1989 that the first Block II satellites were launched. By the summer of 1993, the US launched their 24th Navstar satellite into orbit, which completed the modern GPS constellation of satellites - a network of 24 - familiar now as the Global Positioning System, or GPS. 21 of the constellation of satellites were active at any one time; the other 3 satellites were spares; in 1995 it was declared fully operational. Today's GPS network has around 30 active satellites in the GPS constellation.
Today, GPS is used for dozens of navigation applications, route finding for drivers, map-making, earthquake research, climate studies, and an outdoor treasure-hunting game known as geocaching.

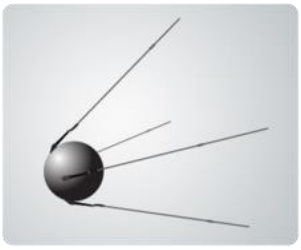


Fig 1.1 – Tracking device shape at that times.

HOW DOES GPS WORK?
There are three parts to a GPS system: a constellation of between 24 and 32 solar-powered satellites orbiting the earth in orbits at an altitude of approximately 20000 kilometers, a master control station and four control and monitoring stations (on Hawaii, Ascension Islands, Diego Garcia and Kawajale) and GPS receivers such as the one in a car.

Each of the satellites is in an orbit that allows a receiver to detect at least four of the operational satellites. The satellites send out microwave signals to a receiver where the built-in computer uses these signals to work out your precise distance from each of the four satellites and then triangulates your exact position on the planet to the nearest few meters based on these distances.

In fact, signals from just three satellites are needed to carry out this trilateration process; the calculation of your position on earth based on your distance from three satellites. The signal from the fourth satellite is redundant and is used to confirm the results of the initial calculation. If the position calculated from distances to satellites "A-B-C" do not match the calculation based on "A-B-D" then other combinations are tested until a consistent result is obtained.

The process of measuring the distance from satellite to GPS receiver is based on timed signals. For example, at 16h45m precisely, the satellite may begin broadcasting its signal. The GPS receiver will also begin running the same random sequence at 16h45m local time, but does not broadcast the sequence. When the receiver picks up the signal from the different satellites, there will be a time lag, because the microwaves take a fraction of a second to travel from the satellite to the receiver. The time lag is easily converted into distance to each satellite. The slight difference between signals from each satellite is then used to calculate the receiver's position.

HOW IT WORKS ON MOBILE DEVICES:

Using GPS on your cell phone depends on the phone. Some phones use GPS only for 911 emergency calls. Sophisticated phones are able to receive and display location maps to show roads and more, and to give directions. These turn-by-turn programs are mostly JAVA-based and work with your cell provider's database.

COMMON USES FOR GPS IN PHONE:

- **Location Tracking**: Some employers use GPS-enabled phones to track their employees' locations, and some business offer location tracking services for GPS-enabled phones. The Wherifone locator phone provides GPS coordinates and can dial emergency phone numbers. Parents and caregivers can track the phone's location by phone or online and can receive notification if it leaves a designated "safe area." Wearable Environmental Information Networks of Japan has also introduced the Dog @ Watch, a GPS watch phone for children.
- **Turn-by-Turn Directions**: GPS-enabled phones with view screens can often display turn-by-turn directions as well as announce them through the phone's speaker. In general, companies that offer these services charge a monthly fee and use a database of maps to provide the directions. The services are only as good as their database -- outdated maps can provide inaccurate directions. Some turn-by-turn direction services include: TeleNav ViaMoto MapQuest Find Mesmart2Go, which requires a separate Bluetooth GPS receiver and a memory card Destinator SP, which is a software package for smartphones.
- **Outdoor Location Services**: Trimble outdoors offers maps and location-based services for hiking, mountain biking, geocaching and other outdoor activities.
- **Other Location-Based Services**: Some companies hope to deliver news, coupons, advertisements and other information to cell phone users based on their location.

GPS DEVICES:

GPS devices are used in various products some of them are as under:-

- Personal Tracker
- Car Gps
- Gps Devices
- Gps Tracker
- Gps Tracking System
- Portable Gps Device
- Gps Tracking Device
- Portable Gps System
- Car Navigation System
- Car Gps System
- Gps Modules
- Gps Navigation System
- Gps Receiver
- Active Gps Antenna
- Aviation Navigation Lights
- Portable Gps Tracker
- Waterproof Gps
- Gps Software
- Bluetooth Gps Receiver
- Gps Antenna
- Portable Gps Receiver
- Handheld Gps
- Gps Vehicle Tracking System
- Gnss Receiver

COMPANIES DEALING WITH GPS DEVICES:

There are various foreign companies too supporting GPS trading but some indian companies are also supporting its trade which are as under:-

- C.S. Aerotherm Pvt. Ltd.
- Global Tele Communication
- Atlas Comnet Pvt. Ltd.
- Link International
- Wift Technologies
- Applied Electro Magnetics Pvt. Ltd.

- ❖ Luxa Control Systems Pvt. Ltd.
- ❖ Sathya
- ❖ G. B. G. Engineering Tools
- ❖ Dvl Groups Ltd.
- ❖ Unitrack
- ❖ Opalsys It Solutions Pvt. Ltd

PRICE OF GPS DEVICES:

The price of GPS devices varies according to company and some of them are as under:-

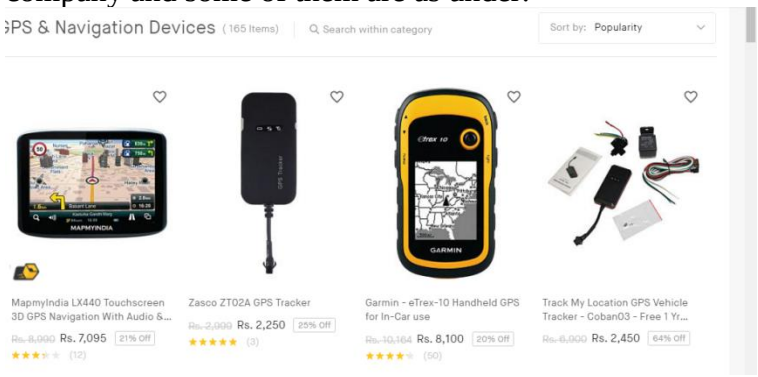


Fig 1.2 – Tracking devices price.

And various sites such as Amazon , snapdeal , flipkart promote selling purchasing of GPS devices.

ACKNOWLEDGEMENTS

Hereby we would like to thank Ms. Kanika Jethwani for their help, right guidance and motivation which they provided us right from the beginning to the end.

TRACKING DEVICES USED IN OTHER AREAS:

Tracking devices are used now a days in other areas, previously it was just to track shipment , in military to search enemy location except these now it is used commercially such as in booking taxi , food and reaching places if you don't know the road to reach a place.

Some example of such services are:-

- OLA Cabs , UBER, Meru Cabs, etc for cab services door to door or plce to place.
- Foodpanda, SWIGGY, Fasos, etc for booking food for lunch , dinner , snacks, breakfasts, etc.
- Hospital using GPS to reach to people / victims for providing support to reach hospital through ambulance.
- Google Maps to reach to places through small routes through car , metro , on foot, etc.
- Social apps such as Facebook, Snapchat , Instagram to tag location to share memories in social media.
- Use in games such as Pokemon GO, Minigami, etc. to provide open world gaming experience.

REFERENCES

1. *"Tracking a suspect by any mobile phone: Tracking SIM and handset". BBC News. 2005-08-03. Retrieved 2010-01-02.*
2. ^ Jump up to:[a] [b] "Location Based Services for Mobiles: Technologies and Standards", Shu Wang, Jungwon Min and Byung K. Yi, IEEE International Conference on Communication (ICC) 2008, Beijing, China
3. **Jump up^** Mobile Positioning Using Wireless Networks
4. **Jump up^** Handset-based mobile phone tracking app example 1: MobileTrack
5. **Jump up^** *Ibrahim, M.; Youssef, M. (2012-01-01).*