International Journal of Enhanced Research in Science Technology & Engineering, ISSN: 2319-7463
Vol. 3 Issue 5, May-2014, pp: (194-197), Impact Factor: 1.252, Available online at: www.erpublications.com

Security of Cloud from Data Mining based Attacks

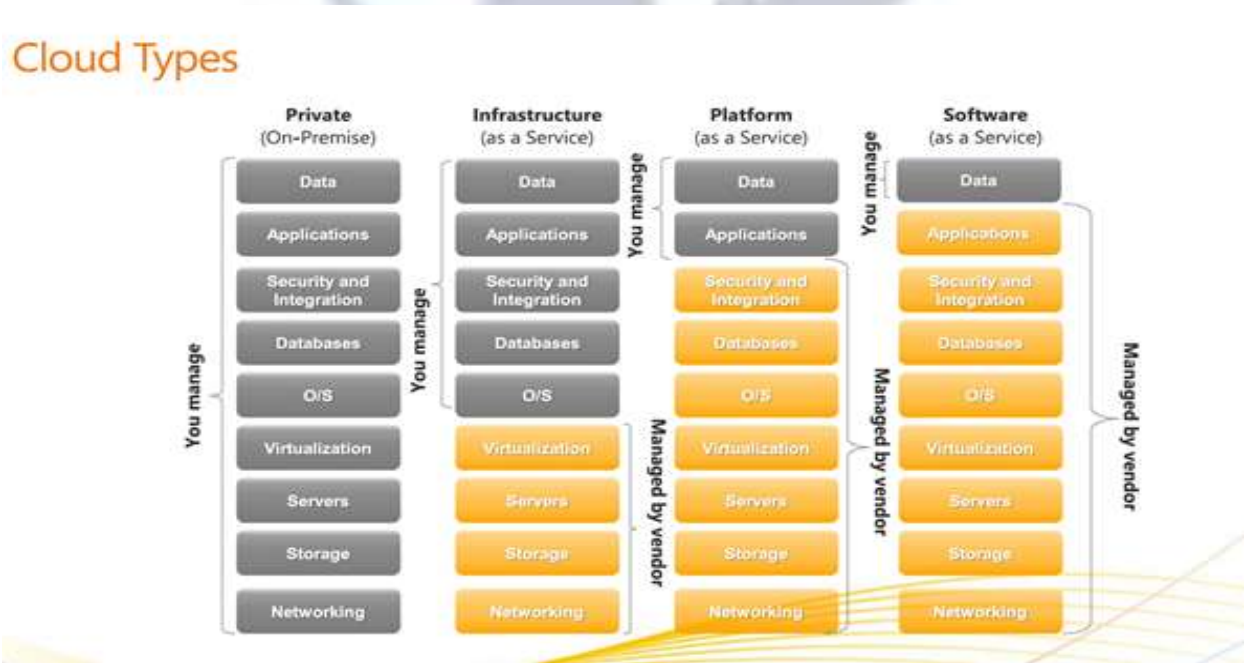
Inderjit Kaur
Department of Computer Science, Lovely Professional University, Phagwara, Punjab, India

Abstract: Data mining is as an essential component in IT and business domain as it expands their current abilities. The objective is to gain the understanding of the project and business requirements, and then converting this knowledge into a data mining problem definition and then a plan to achieve those objectives. The current needs of IT sector make the Cloud computing come into existence. There is a growing trend of using cloud environments for storage and data processing needs. Cloud computing provides means to improve or add abilities on- demand without making any investments in infrastructure set up or training new employees. Reasoning processing may involve any subscription-based or pay-per-use support which is provided Online immediately which expands all horizons. But people still have concerns over security and privacy of data which is uploaded on the cloud. These concerns are much more severe in case of public clouds as they are more vulnerable to different security breaches.

Keyword: - Data mining, Cloud computing and Security.

1. INTRODUCTION

We live in a period of enormous information; its increasing amount has inserted a colossal potential and expanded complexity and risks such as data security as well as irrelevance and information overload. This information is stored in huge warehouses. Data or information mining is an alternate enhancement to help ventures to keep tabs on data in their warehouses. Cloud computing is an Internet-based computing, whereby shared resources, software, and information are provided to computers and other devices on demand. It is an upcoming paradigm that offers tremendous advantages in economical aspects, such as reduced time to market, flexible computing capabilities, and limitless computing power. Popularity of cloud computing is increasing day by day in distributed computing environment. There is a growing trend of using cloud environments for storage and data processing needs. To use the full potential of cloud computing, data is transferred, processed, retrieved and stored by external cloud providers. However, data owners are very sceptical to place their data outside their own control sphere. Their main concerns are the confidentiality, integrity, security and methods of mining the data from the cloud. Security becomes a major concern when using public cloud services.



2. RESEARCH AREAS in CLOUD

Most research in cloud is done on security issues in cloud.

- **Cloud security:** Cloud security principles are the same that apply to on-site computing. Security issues are same but complexity is more because of large number of users and data.
- **Identity management.** Managing access to computer applications, data, softwares, resources, services by checking identity information of user properly.
- **Detection and forensics.** Separating legal activity from illegal activities. This helps to keep away malicious users and attacks.
- **Encryption.** Using code to protect sensitive information in data warehouses and other assets. There is a lot of scope in this area as security is the main concern in cloud.
- **Cloud governance and compliance:** Generally, governance defines who is responsible for what and what policies and procedures people or groups need to follow. Cloud governance requires that you are responsible for your own infrastructure as well as that infrastructure which you don't have total control over. There are two key components: understanding compliance that is acting according to certain standards and performance goals in business.
- **Data in the cloud:** Data security and privacy is required in cloud to manage data, it includes all control over moving data from one point to another. It also includes managing the resources and data storage for large-scale processing of data.
- **Data mining in cloud:** It is the process which involves extracting meaningful information from scattered or unstructured internet sources of data. Secure and efficient services can be provided to users when organizations allow centralization of the management software and storage of data.
- **Cloud manageability:** A consistent view for management is required for both cloud based and on-premises service environments. Manageability as the name states; manages the resources/assets and the quality of service (QOS) which is received from any cloud service provider. Many other areas like defence, business operations, and health and human services can also use cloud quality services efficiently.

3. DATA MINING in CLOUD

Extracting hidden patterns or information which can be very valuable from large data centres is data mining. This technology helps companies to keep their focus on crucial information in their data warehouses which were otherwise neglected. The tools used for mining make future predictions and trend behaviours by enabling businesses to control situations even before they occur by making proactive decisions based on knowledge discovery. The automatic analyses of data provided by data mining tools now-a-days can easily be declared better than the analyses of past events done by decision support systems. Traditional methods are not very effective now as the size and complexity of datasets has increased. Automatic data processing which can also be indirect has been added in data analysis. Some other inventions in computer science field like cluster analysis, neural networks, genetic algorithms (1950s), decision trees (1960s) and support vector machines (1990s) also added to this. When these methods are applied to data sets to find hidden patterns or forecasting, it is referred to as data mining.

Raw data is collected and transitioned into valuable data. To sort through this raw data and to identify the forms or patterns and then establishing relationships between them is data mining. Different parameters for data mining which discover different patterns include association, classification, clustering, forecasting, sequence or path analysis. Data mining has multiple applications in real world such as Hospital, Student Management, Airline Reservation, Forecasting, Biometrics, Mathematics, Geographical, Web Mining, Parallel Processing, Data Integrity, etc. Information mining systems and requisitions are sincerely needed in the distributed computing ideal model. As distributed computing is entering all the more in all degrees of business and experimental processing, it transforms into an incredible issue to be concerned by information mining. The information mining in Cloud Computing grants associations to make the incorporated administration of programming and information space, guaranteeing the effective, solid and secure administrations for their clients.

4. "DMCLOUD"

The cloud-based services in the Microsoft suite include a new technical preview of Data Mining in the Cloud called "DMCloud". Some basic data mining tasks can be performed with cloud-based connection called Analysis Services.

DMCloud is valuable capability for IWs (Information Warfare) as it would use SQL Server Data Mining without the additional burden of Analysis Services. As all these services require a technology professional to install it. The Information Warfare would just need an internet connection to use the DMCloud services irrespective of their location on this globe. The old Excel Data Mining add-in had some Table Analysis Tools; these tools can also be used with DMCloud today.

Some data mining tasks which can be done using these tools include:

- Key Influences can be analyzed.
- Detection of Categories
- Fill From Example
- Forecasting
- Highlighting Exception
- Analysis of scenario in given data
- Calculation of Predictions
- Shopping Basket Analysis

Data mining in cloud computing is to deduce or extract useful information from scattered or unstructured sources of data from web. It is applicable in various fields like student management, hospitals, scientific observations and various other places. Services which are secure and efficient can be provided to users of any organization if the storage of data and management of software is centralized in that organization. It can easily be found that how some data mining tools like SaaS, PaaS and IaaS are used in cloud to extract useful information. It is used in business field for almost all types of research like marketing, technical and patent research. Any cloud can be a provider for natural processing system for languages. Some leading cloud computing providers are Amazon Web Services, OpenStack, Windows Azure.
Data mining tools search in forums about different topics to build lists and other information is obtained. These services are also used by companies to check what type of reviews and news are floating in the internet world regarding services and products provided by them.

Then the data they get helps these companies to take action about their products as required. An information retrieval practical model has been proposed which uses data mining in a cloud computing environment through a multi-agent system. It is important to protect the data centers because cloud is all about virtualization and sharing of resources. So it is important to know which user has access to which service. Data mining algorithms and their various applications make the work of multi-agent systems a lot easier to retrieve meaningful information. It should also be kept in mind by the users that the requests they make to the data warehouse should also be simple and clear. When data mining tools are used on large data sets in cloud for a long time, it has positive as well as negative effects. Positive effects include giving warnings and other forecasting about future trends and would be problems. Negative effects are analyzing some user's private data which can lead to misuse of somebody's private data.

5. RESEARCH OBJECTIVES

1) To implement cloud security for data mining
Data security is the major concern in public clouds. So in this paper the cloud is hosted publically.

2) Finding the loop holes or the threats related to data mining

a) Uploading the data on the single cloud
This can lead to data loss in case of events like network outage, the cloud provider going out of business, malware attack etc. As the entire data belonging to a client remain under a single cloud provider, both inside and outside attackers (inside attacker like malicious employees and outside attacker like hackers) get the benefit of using data mining to a great extent.

b) Uploading data on distributed cloud providers
Although this scenario will protect the client's data as the data will be distributed to different cloud providers. But it will increase the cost to the client as purchasing different cloud will increase the cost.

3) To enhance security in cloud systems by creating user access policies.
We will be enhancing the security by using single cloud provider and dividing single cloud into different zones thereby saving a cost of the client and also enhancing the security. We will segregate data by creating virtual partitions of data for saving and allowing user to access data in his partition only. Each user will have the rights according to the role of the client i.e. role based access polices. Use of virtual partition and enhanced user access control in cloud system will improve data security and thereby fixing the threats in data mining to Personal/private data in cloud systems.
Enhanced Cloud system will be compared with existing secure cloud systems. We will compare enhanced system against security, performance & ease of use.

6. RESEARCH PLAN AND METHDOLOGY

Cloud computing is an Internet-based computing, whereby shared resources, software, and information are provided to computers and other devices on demand. Upcoming cloud paradigm will offer tremendous advantages in economical aspects, such as reduced time to market, flexible computing capabilities, and limitless computing power. Distributed computing field has increasing demand of cloud computing day by day. There is a growing trend of using cloud environments for storage and data processing needs. To use the full potential of cloud computing, data is transferred, processed, retrieved and stored by external cloud providers. We will be implementing cloud security aspects for data mining by implementing cloud system. After implementing cloud infrastructure for data mining for cloud system we shall be evaluating security measure for data mining in cloud. We will be fixing threats in data mining to Personal/private data in cloud systems. Cloud based systems saves data off multiple organizations on shared hardware systems. Data segregation is done by encrypting data of users, but encryption is not complete solution. We can do segregate data by creating virtual partitions of data for saving and allowing user to access data in his partition only. Malicious activity monitoring is a tough task in cloud system as logging data might be spread over multiple hosts and data centres. Restricting user to its own virtual partition only will not allow logs to be dispersed allowing access to logs for monitoring easily. User access is another major concern in restricting user access is a major challenge in cloud based storage system. Use of virtual partition and enhanced user access control in cloud system will allow us to improve data security. The tool to be used for implementation is Eclipse IDE.

7. FUTURE SCOPE

It is already stated earlier in this report that data mining is an approach to find patterns in databases. So a basically this research is to classify networks and cloud assets and on the topic of privacy, confidentiality and integrity. Many data mining models and other security concerns are also investigated. In this dissertation, basic focus is on the data in cloud, its privacy, mining and confidentiality, as these are where the major security concerns are at this moment.

Research on classification of data in cloud has already been extensively done; so this dissertation will use the results of these researches and analyze the security requirements which are important for keeping data secure. Relying on cloud computing millions of users store their data on a cloud which possess lot many cloud storage risks like unauthorized access, data loss etc. Privacy of data is a major concern in people who use public cloud services, so an approach is proposed to keep data safe and secure also keeping sure only authorized personnel can access data.

REFRENCES

[1]. Parikshit Prasad, Badrinath Ojha, Rajeev Ranjan shahi, Ratan Lal, “3 Dimensional Security in Cloud Computing”, IEEE, 2011
[2]. Uma Somani, Kanika Lakhani, Manish Mundra, “Implementing Digital Signature with RSA Encryption to Enhance Data Security of Cloud in Cloud Computing”, IEEE, 2010.
[3]. Fadi Aloul, Syed Zahidi, Wassim El-Hajj, “Two Factor Authentication Using Mobile Phones”, 2008.
[4]. Balachandran Reddy, Cloud computing security issues and challenges, 2009
[5]. Special Publications 800-145 “National Institute of Standard and Technology (NIST)” http://en.wikipedia.org/wiki/Cloud_computing.
[6]. http://cloudcomputing.sys-con.com/node/1744132.
[7]. Zhidong Shen, Qiang Tong, “The Security of Cloud Computing System enabled by Trusted Computing Technology”, 2010.
[8]. http://www.cloudcomputingchina.cn/Article/ /200909/306.html.
[9]. Pekka Riikonen, “RSA Algorithm”, 2002 [11] Torry Harris, “Cloud Computing An Overview”.