CLOUD COMPUTING USING DATA MINING

Khyati Chaudhary

Abstract— **Cloud can be meant as an infrastructure that provides resources and/or service over the internet. Cloud computing has demonstrated that processing very large datasets over commodity clusters can be done by giving the accurate programming model. Data mining is the procedure of analyzing data from different perspectives and briefs it into constructive information. Data mining has been measured as a necessary component in business domain. Cloud computing has revolutionized the way computing and software services are delivered to the clients on demand. The integration of data mining techniques into normal day-to-day activities has become common place. Using the single cloud provider is a major problem among the clients in the cloud as the outside attacker can analyze their data for a long time to gain the sensitive information. Cloud computing provides means to improve or add abilities on-demand without making an investment in setting up infrastructures, training new employees. Therefore, security concerns among users of the cloud have become a major barrier to the widespread growth of cloud computing. Hence, this is a big concern for many clients of cloud. In this paper, we identify the data mining based privacy risks on cloud data and we have given the data mining based attacks on cloud data as well as describe how data mining is used in cloud computing.**

Index Terms— **Cloud, Cloud computing, Data mining, databases privacy and security.**

I. INTRODUCTION

Data Mining is used for extracting potentially valuable information from rare data. Every day people are confront with targeted advertising and data mining techniques help businesses to become more efficient by reducing costs. Data mining allows users to analyze data from diverse dimensions and categorize it and summarize the relationships identified. The implementation of data mining techniques through cloud computing will allow the users to regain meaningful information from virtually integrated data warehouse that reduces the costs of infrastructure and storage. Cloud computing entrusts remote services with a user's data, software and computation. A cloud can be a storage cloud that provides file based storage service. Cloud computing provides a good model for the providers to deploy the computing infrastructure and applications on-demand. Cloud computing offers better flexibility to users by connecting to a variety of computing resources and allowing access to IT enabled services but it has the risk of privacy of user data and security. Hence security among the users of cloud is the most important concern. One of the security concern in cloud

computing is data mining based attacks which involves that the data can be analyzed continuously by an unanonymous person to get the valuable information. High performance data mining system is designed for taking benefit of powerful and shared pools of processors. In this case, data is distributed over the processors and the computation is done using message passing concept. After that all the computation results are gathered and this process is repeated on the new data on the processor. PC researchers, economists, Investigators, mathematicians, sociologists and different researchers are looking for access to the huge amount of information to extract meaningful information and knowledge Cloud is not only a source of enormous static data but also a provider of high processing ability at low cost. Cloud has been giving providers chance to analyze user data for a long time. A report stated that 90 percent of corporate information, incorporating archives, website pages and email is unstructured.

Confidentiality of user data in the cloud is another huge concern. Outside attackers who manage to get access to the cloud can also analyze data and break user privacy. Information mining is an alternate enhancement extraction of concealed prescient data from gigantic databases. The information mining devices expect future examples or allowing associations to make proactive and learning driven choices. Purpose of information mining is to find at one time fuzzy relationships around the information, especially when the information starts from dissimilar databases. Cloud computing facilitate end-users or small companies to use computational resources such as software, storage.

Recent trends of data analysis involve mining which is closely associated with statistical analysis of data. Data mining can be a potential threat to cloud security considering the fact that whole data belonging to a particular user is stored in a single cloud provider. Several analysis techniques are being used by cloud service providers. For instance, Google uses data analysis techniques to analyze user behaviours and recommend search results. Attackers can use few of these techniques to extract valuable information from the cloud. The single storage provider approach gives the provider opportunity to use powerful mining algorithms that can extract private information of the user. As mining algorithms require a logical amount of data, the single provider architecture suits the reason of the attackers. This single cloud storage provider approach also eases the job of attackers who have unauthorized access to the cloud and use data mining to extract information. Therefore, the privacy of data in the cloud has become a major concern in recent years.

Categorization allows to recognize sensitive data and to take appropriate initiatives to maintain privacy of such data. Fragmentation and distribution of data among providers reduce the quantity of data to a particular provider and consequently minimize the risk linked with information leakage by any provider. This distribution is done according

to the sensitivity of data and the reliability of cloud providers. A cloud provider is given a particular data chunk only if the provider is reliable enough to store chunks of such sensitivity. Distribution restricts an attacker from having access to a plenty number of chunks of data and hence prevents successful extraction of valuable information via mining.

Cloud computing is popular due to its mobility, huge availability and low cost. Alternatively, it bring more threats to the security of the company's data and information. Cloud is an infrastructure that provides on-demand resources or services over the Internet usually at the scale and reliability of a data centre. A structure which has storage cloud; provides storage services (block or file-based services); a data cloud provides data management services (record-based, column based or object based services) and a compute cloud provides computational services. Often these are stacked together to serve as a computing platform for developing cloud-based applications. Cloud computing services assure users are able to store their data in servers and access their data from anywhere and they need not worry about the loss of data due to disk faults, system breakdown.

II. BACKGROUND

Internet is becoming a primary tool in our everyday life both professional and personal as its users are becoming more frequent. The emerging cloud computing trends provides the sole benefit of unprecedented access to valuable data for its users. The expansion of the World Wide Web (www) has resulted in a big amount of data that is now freely available for user access. Several data mining methods are used to discover the hidden information in the Web. However, Web mining does not only mean applying data mining techniques to the data stored in the Web.

Cloud is an infrastructure that consists of services delivered through shared data centres and appearing as a single point of access for consumers computing needs. Cloud computing also provides demanded resources over the internet. Cloud computing can also be a data cloud that provides record-based, column-based or object-based services. These all type of clouds are set up as a stack of cloud services that provide computing platform to develop cloud based applications. The cloud computing is basically the assets and administrations required to perform capacities with alterably evolving requirements.

III. DATA MINING PARAMETERS

Data mining parameters include:

1. Association- Look for patterns where one event is connected to another event

2. Sequence or path analysis- Look for patterns where one event leads to another later event

3. Classification- Look for new patterns

4. Clustering - Finding and visually documenting groups of facts not known earlier.

5. Forecasting - Discovering patterns in data that can lead to logical predictions about the future. This area of data mining is known as predictive analytics.

IV. HIGHLIGHTS ON CLOUD SERVICE

Cloud computing services includeInfrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). There are three types of cloud services as follows-

Infrastructure as a Service(IaaS)-

(i) Delivers computer infrastructure as a utility service, typically in a virtualized environment.

(ii) Provides enormous potential for extensibility and scale.

Platform as a Service (PaaS)-

Delivers a platform or solution stack on a cloud infrastructure. Sits on a top of the IaaS architecture and integrates with development and middleware capabilities as well as database, messaging and queuing functions.

Software as a Service (SaaS)-

(i) Delivers the application over the Internet or Intranet via a cloud Infrastructure.

(ii) Built on underlying IaaS and PaaS Layer.

```
┌─────────────────────────────────────────┐
│             CLOUD CLIENTS                │
│  Web browser, mobile app, thin client,   │
│         terminal emulator...             │
└─────────────────────────────────────────┘
                    ⇕
┌─────────────────────────────────────────┐
│ SaaS                                     │
│      CRM, E-MAIL, VIRTUAL DESKTOP....     │
├─────────────────────────────────────────┤
│ PaaS                                     │
│      DATABASE, WEB SERVERS.....          │
├─────────────────────────────────────────┤
│ IaaS                                     │
│      VIRTUAL MACHINES, SERVERS.....      │
└─────────────────────────────────────────┘
```
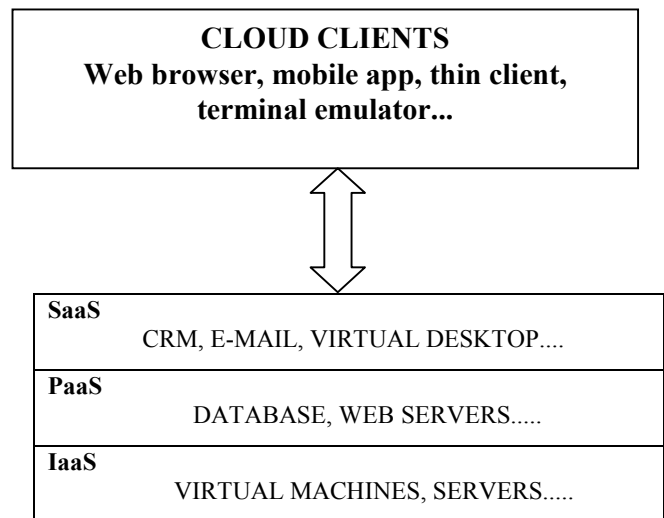
FIGURE 1: Layers of Cloud Computing

Big corporate such as Amazon, Google and Microsoft are providing cloud services in a variety of forms. Google provides Platform as a Service (PaaS) which is known as Google App Engine (GAE) and facilitates hosting of web applications. Although cloud computing is a major means of achieving high storage and computing services at a low cost. Cloud has several security issues involving assurance and confidentiality of data. A user entrusting a cloud provider may lose access to his data temporarily or permanently due to an improbable incident such as a malware attack or network outage. Such an unlikely event can do significant harm to the users.

V. CONNECTION BETWEEN CLOUD COMPUTING AND DATA MINING

Data Mining is the main growing field in IT industry which is also known as Knowledge Discovery in Databases (KDD). It is used to discover patterns from large volumes of data. In data mining, the main areas are like Frequent Pattern Mining, Association Rule Mining etc. Cloud Computing and Data Mining are interrelated to each other and having its advantages and disadvantages. The advantage is: data mining

has been used by cloud providers to provide improved service to clients. The disadvantage is: attackers outside the cloud provider, who is not having authorized access to cloud, will also use data mining to extract data from cloud. The extraction of useful data from cloud involves 2 factors: suitable amount of data and appropriate mining algorithms. There are so many mining algorithms which will work good to extract useful information from cloud which violated the user data privacy.

VI. DATA MINING IN THE CLOUD

Data mining is one of the growing fields in computer industry that deals with discovering patterns from large data sets. Mining is preferably used for a huge amount of data and related algorithms often require large data sets to create quality models. According to the survey done by Rexer Analytics, 7% data miners use cloud to analyze data. As cloud is a enormous source of centralized data, data mining gives attackers a great advantage in extracting valuable information and thus violating clients' data privacy. The importance of client privacy is a tentative issue as all clients do not have the same demands regarding privacy. Some are satisfied with the current policy while others are quite concerned about their privacy. If the client itself is a company providing services to others, the violation of privacy of the client affects the privacy of its customers.

The information mining in Cloud Computing grant associations to make the integrated administration of programming and information space, guarantee the effective and secure administrations for their clients. The Microsoft suite of cloud based administrations presents specialized steal of Data Mining in the Cloud. The data mining tasks include:

1. Analyze Key Influencers
2. Detect Categories
3. Fill From Example
4. Forecast
5. Highlight Exceptions
6. Scenario Analysis

A. Data Mining: A Potential Threat to Privacy

The extraction of useful information using data mining depends on two main factors: (a) proper amount of data and (b) appropriate mining algorithm. A number of mining algorithms are enough to extract information up to the limit that violates client privacy.

Clustering algorithms can be used to classify people or entities and are suitable for finding behavioural patterns. Association rule mining can be used to discover association relationships among large number of business transaction records. Thus analysis of data can reveal private information about a user and leaking this sort of information may do major harm. As a result, data mining is becoming more powerful and possessing more threat to cloud users. For instance, multivariate analysis identifies relationship among variables and this technique can be used to determine the financial condition of an individual from his/her buy-sell records. The data mining in Cloud Computing allows organizations to centralize the management of software and data storage with guarantee of efficient, reliable and secure services for their users." The main effects of data mining tools being delivered by the cloud are: the customer only pays for the data mining tools that he/she needs; that reduce the costs.

Data mining in the cloud computing is a process of extracting structured information from unstructured or semi-structured web data sources. The data mining in cloud computing allows organizations to centralize the management of software and data storage with assurance of efficient, reliable and secure services for their users. The implementation of data mining techniques through cloud computing will allow the users to retrieve meaningful information from virtually integrated data warehouse that not only reduce the cost of infrastructure but also storage.

B. CLOUD COMPUTING DEPLOYMENT MODELS

A cloud computing provides following basic service models:
1) Public Clouds: Public clouds are frequently hosted away from customer site and they provide flexible infrastructure to cut down customer risk and cost.
2) Private Clouds: Private clouds are assembled for deployment of one customer solely, furnishing the most extreme control over information, security, and nature of administration. In this model, an organization can introduce, arrange and work the foundation to help a private cloud inside an organization's undertaking data centre.
3) Hybrid Clouds: Hybrid clouds join the aspects of both open and private cloud models. The capacity to incorporate a private cloud with the assets of an open cloud might be utilize to administer administration levels.
4) Community Clouds: In Community Cloud, the cloud base is imparted by frequent associations.

C. SECURITY FOR CLOUD COMPUTING

Cloud has some security issues concerning affirmation and classification of information. A client entrusting a cloud supplier may lose access to his/her information incidentally or forever because of a doubtful circumstance; for example, a malware trap or system blackout. Such an impossible occurrence can do notable harm to the clients. Security issues have been classified into 2 general classifications: Security issues visage by cloud suppliers and security issues visage by their clients. As a rule, the supplier may verify that their framework is secure in as much as the customer might as well determine that the supplier has taken the correct efforts to establish safety to defend their information.

CLOUD NAME	DESCRIPTION
Sun Microsystems Sun Cloud	More available application than any other open operating system..
IBM Dynamic Infrastructure	Integrated power management to help you predict, monitor, plan, and actively manage power consumption.
Amazon EC2	Intended to make web-scale computing easier for developers.
Google App Engine	No limit of the free trial period if you do not exceed the quota allotted.
Microsoft Azure	Currently offer a "development accelerator" discount plan. 15-30 % discount off consumption charges for first 6 months.

TABLE 1: Presents details on the top cloud computing companies and their products key features

CONCLUSION

Relying on cloud computing, millions of users store their data on a cloud which have many cloud storage risks like unauthorized access, data loss etc. Privacy of data is a major concern in people who use public cloud service. It is proposed to employ cloud security aspect for data mining by implementing cloud system. Threats will be fixed in data mining to personal/private data in cloud systems. Data mining technologies provided through cloud computing is an totally necessary characteristic for today's businesses to make proactive, knowledge driven decisions as it helps them have future trends and behaviours predicted. This paper provides an overview of the necessity and benefit of data mining in cloud computing. As the need for data mining tools is growing every day, the ability of integrating them in cloud computing becomes more severe.

REFERENCES

[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] http://en.wikipedia.org/wiki/Cloud_computing Kaur et al., International Journal of Advanced Research in Computer Science and Software Engineering 4(3), March - 2014, pp. 1178-1183 © 2014, IJARCSSE All Rights Reserved Page | 1183

[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] [Sanjay Ghemawat, Howard Gobioff, and Shun-Tak Leung. "The Google File System". In SOSP, 2003.

[10] M. Kantardzic. Data Mining: Concepts, Models, Methods and Algorithms.John Wiley & Sons, Inc., 2002.