International Journal of Advanced Research in Engineering and Applied Sciences

ISSN: 2278-6252
Impact Factor: 7.358

MITIGATING THE DATA MINING BASED ATTACKS ON CLOUD BY DEVELOPING THE SECURED MECHANISM

Surbhi Singh, Research Scholar, Department of Computer Science & Engineering, Deenbandhu Chhotu Ram University of Science & Technology, Sonipat

Abstract: *Cloud computing is commonly used with data mining techniques. Security is still the major concern in case of databases in cloud or DataCloud. Major privacy issues arise in public clouds since they are open to all for services. These privacy issues can be solved according to me by creating user access policies. Data mining is important for business field also. It has focus on understanding the objectives of the project and requirements of business, and then converting this knowledge into a data mining problem definition. There are security issues related to data mining attacks in cloud computing. Form a preliminary plan to design and achieve the security objectives. But Cloud computing comes into perspective only when you think about the needs of IT sector: a way to improve potential or add abilities on the fly without making an investment in new facilities, coaching new employees, or certification new application. Reasoning processing involves any subscription-based or pay-per-use support that, immediately over the Online, expands IT's current abilities.*

Keywords: *Cloud Computing, Data Mining, Confidentiality, privacy, threats.*

1. INTRODUCTION

Cloud computing is a concept for large number of computers connected with each other so that hosted services can be delivered over the Internet. Clouds can be private, public for use, community or a combination of any of these. Various associations are picking as an elective to building their specific IT base to have databases or programming so the association may have entry to its data and customizing over the Internet. The use of Cloud Computing is getting conspicuousness due to its transportability, enormous approachability and insignificant expense. It joins matrix registering, utility processing, virtualization, and software bunching and so on. Distributed computing blankets a share of the thoughts of conveyed, framework and utility registering. Cloud is by and large a virtualization of assets that administers and oversees itself. The cloud computing is fundamentally entering the assets and administrations required to perform capacities with alterably evolving

requirements. The administration engineer appeals access from the cloud instead of a particular endpoint or named asset.

1.1 DATA MINING IN CLOUD

Extracting hidden patterns or information which can be very valuable from large data centers 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 behaviors 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 patters 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 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.

1.1.1 Security of Cloud from Data Mining

Cloud has a few security issues concerning affirmation and classification of information. A client entrusting a cloud supplier may lose access to his information incidentally or forever because of a doubtful occasion, for example, a malware ambush or system blackout. Such an impossible occasion can do noteworthy harm to the clients. Secrecy of client information in the cloud is a huge concern. There is a wide mixed bag of security issues identified with distributed computing however these issues have been classified into 2 general classifications: Security issues visaged by cloud suppliers and security issues visaged by their clients.

As a rule, the supplier may as well verify that their framework is secure inasmuch as the customer might as well determine that the supplier has taken the right efforts to establish safety to defend their information. Distributed computing could display diverse dangers to an organization than old IT results. Cloud security contemplations are characterized into any mixed bag of extents and these sizes are aggregative into 3 general zones: Security and Privacy, Compliance, and Legal or composed understanding issues.

Data security is not limited to securing it in cloud but also providing security methods while the data is on move, being transferred. The existing encryption methods used by clouds are

not anymore enough because encryption keys are not known to the customers, and are managed off-cloud which can cause problems. A lot of data mining based attacks can be helpful for intruders to derive useful and sensitive information. With the availability of so many data mining tools and algorithms, it is now relatively easy to mine information from data centres in clouds, mostly public clouds. My implementation process first shows how it is possible to apply mining on files available in the cloud by the service provider.

To secure data from the cloud service provider the user (here admin.) uses the proposed model, according to which virtual partitions are created in the cloud. Each file which is uploaded is divided into multiple small segments and segregated in these virtual partitions. Breaking down the file in segments which are distributed in the partitions helps to thwart any mining based attack and forms an effective security technique.

2. PROPOSED SOLUTION TO PREVENT DATA MINING ATTACKS ON CLOUD

Distributing data affects all mining algorithms, like prediction algorithm also gives misleading results because numbers of observations are less. Segregation of data reduces the number of available data and affects the results. It changes the result of clustering algorithm as well as the clusters and cluster heads move away from the original ones. Regression analysis also requires many variables which in turn require number of sample cases, which is obviously not available. To make data more secure public cloud service providers can make user access policies, which will enable only authorized users to access their own data, and not by anyone else.

It is proposed to implement cloud security aspects for data mining by implementing cloud system. After implementing cloud infrastructure for data mining for cloud system, security measure for data mining in cloud will be evaluated. Threats will be fixed 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. Data can be segregated 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 users to their own virtual partition will not only 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. Enhanced Cloud system will be compared with existing secure cloud systems. Then this enhanced system with security will be compared with performance & ease of use.

3. CONCLUSION

In this paper, the impact of data mining on Cloud is discussed and a distributed structure to eliminate mining based privacy threat on cloud data is introduced. The proposed approach combines categorization, fragmentation and distribution, and prevents data mining by maintaining authentication, splitting data into segments and storing these storing this data into virtual zones. The proposed system provides an effective way to protect privacy from mining based attacks; it also overcomes the shortcomings of the previous model. In the earlier approach where data is distributed to various cloud service providers, which degrades system performance, for example if the client needs to access all data frequently then there is performance overhead. Centralizing data instead of distributing it to multiple cloud service providers also saves cost and is time- efficient.

Our approach combining categorization, fragmentation and distribution, prevents data mining by maintaining privacy levels, splitting data into chunks and storing these chunks of data to appropriate cloud providers.

Although the proposed system provides an effective way to protect privacy from mining based attacks, it introduces performance overhead when client needs to access all data frequently, e.g. client needs to perform a global data analysis on all data. The analysis may have to access data from multiple locations, with a degraded performance. In future, we look forward to improve our system by reducing such overhead.

REFERENCES

[1] Alawode A. Olaide (2013) "*On Modelling Confidentiality Archetype and Data Mining in Cloud Computing*", Dept of Computer Systems Engineering**.,** Imo State University, Oweri, Nigeria.

[2] Aloul Fadi, Zahidi Syed and El-Hajj Wassim (2008) "*Two Factor Authentication Using Mobile Phones*".

[3] Ambulkar Bhagyashree and Borkar Vaishali (2012) "*Data Mining in Cloud*

Computing".

[4] Benny Pinkas, Ahmad-Reza Sadeghi, and Nigel P. Smart (2012), "Secure Computing in the Cloud, Dagstuhl Reports, Daghstuhl, Germany.

[5] Dubitzky Werner (2008) Data Mining Techniques in Grid Computing Environments, Wiley-Blackwell John Wiley & Sons, West Sussex, UK.

[6] Himel Dev Himel, Tanmoy Sen. Tanmoy, Basak Madhusudan and Ali Mohammed Eunus (2012) "*An Approach to Protect the Privacy of Cloud Data from Data Mining Based Attacks*" Department of CSE, Bangladesh University of Engineering and Technology (BUET), Dhaka.

[7] Hsu Wen-Feng, Yuan Shyan-Ming, Luo Guo-Heng and Tsai Ching-Tsorng (2012) "*Constructing Private Cloud Storage Using Network Attatched Storage*" 9th International Conference on Ubiquitous Intelligence and Computing and 9th International Conference on Autonomic and Trusted Computing, Taiwan.

[8] http://en.wikipedia.org/wiki/Cloud_computing.

[9] http://www.cloudcomputingchina.cn/Article/ /200909/306.html

[10] http://www.examiner.com/article/five-advantages-and-disadvantages-of-cloud-computing.

[11] Hurwitz Judith, Bloor Robin, Kaufman Marcia, Halper Fern (2010) Cloud Computing for Dummies, Wiley Publishing, Inc., Indianapolis, Indiana.

[12] Prasad Parikshit, Ojha Badrinath, Shahi Rajeev Ranjan, Lal Ratan, (2011) "*3 Dimensional Security in Cloud Computing*", IEEE.

[13] Ramgovind S, Eloff MM and Smith E (2010) "*The Management of Security in Cloud Computing*" School of Computing, University of South Africa.

[14] Reddy Balachandran (2009) "*Cloud computing security issues and challenges*"

[15] Ruxandra-Ştefania PETRE, *Database Systems Journal* vol. III, no. 3/2012, "Data mining in Cloud Computing", Bucharest Academy of Economic Studies.

[16] Shen Zhidong and Tong Qiang (2010) "*The Security of Cloud Computing System enabled by Trusted Computing Technology*".

[17] Zhou Minqi†, Zhang Rong§, Xie Wei†, Qian Weining† and Zhou Aoying† (2010) "*Security and Privacy in Cloud Computing: A Survey*" Sixth International Conference on Semantics, Knowledge and Grids. †Software Engineering Institute, East China Normal University, Shanghai, China. §National Institute of Information and Communications Technology, Kyoto, Japan.