www.ijecs.in
International Journal Of Engineering And Computer Science ISSN: 2319-7242
Volume 5 Issue 5 May 2016, Page No. 16427-16434

Cloud Computing: "Secured Service Provider for data mining"

Garima Gupta, Dr. Rajesh Pathak

M.Tech Student

HOD (CS Department)

Greater Noida Institute of Technology and Engineering

Abstract

Data safety and procedure in method are the most demanding scrutinize work going on, at present, in cloud computing. This is for the reason that of the clients sending their complex data to the cloud providers for obtaining their facilities. Thus, security concerns among users of the cloud have become a major barrier to the extensive growth of cloud computing. One of the security concerns of cloud is data mining based isolation attacks that involve analyzing data over a long period to extract valuable information. In particular, in current cloud architecture a client entrusts a single cloud provider with his data. It gives the provider and outside attackers having unlawful access to cloud, an occasion of analyze client data over a long period to take out sensitive information that causes privacy violation of clients. Therefore, to rise above these difficulties, a model (cloud architecture) is explained in this paper which accepts only those data which are essential in an encoded form, performs the service opted by the client and sends the result in the encoded format to be implicit by the specific client.

Keywords-cloud computing, data mining, cloud server, web services, SaaS

1. Introduction

Cloud computing facilitates end-users or small companies to use computational resources such as software, storage, and processing capacities belonging to other companies (cloud service providers). Cloud services include Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS) [2]. Big corporate like Amazon, Google and Microsoft are providing cloud services in various forms. Amazon Web Services (AWS) provides cloud services that include Amazon Elastic Compute Cloud (EC2), Simple Queue Service (SQS) and Simple Storage Service (S3) [9]. Google provides Platform as a Service (PaaS) known as Google App Engine (GAE) and

facilitates hosting web applications [6]. Microsoft also provides cloud services in the form of Windows Azure, SQL Azure, Windows Intune etc.

Although cloud computing is a powerful means of achieving high storage and computing services at a low cost, it has not lived up to its reputation. Many potential users and companies yet lack interest in cloud based services [11]. One of the main reasons behind this lack of interest involves security issues. Cloud has several security issues involving assurance and confidentiality of data [6]. A user entrusting a cloud provider may lose access to his data temporarily or permanently due to an unlikely event such as a malware attack or network outage. Confidentiality of user data in the cloud is another big concern. Cloud has been giving providers an opportunity to analyze user data for a long time. In addition, outside attackers who manage to get access to the cloud can also analyze data and violate user privacy. Cloud is not only a source of massive static data, but also a provider of high processing capacity at low cost. This makes cloud more vulnerable as attackers can use the raw processing power of cloud to analyze data [11].Various data analysis techniques are available now-days that successfully extract valuable information from a large volume of data. These analysis techniques are being used by cloud service providers. For example, Google uses data analysis techniques to analyze user behaviors and recommend search results [11]. Attackers can use these techniques to extract valuable information from the cloud. The recent trends of data analysis involve mining which is closely associated with statistical analysis of data [22]. Data mining can be a potential threat to cloud security considering the fact that entire data belonging to a particular user is stored in a single cloud provider. 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 reasonable amount of data, the single provider architecture suits the purpose of the attackers. This approach (single cloud storage provider) also eases the job of attackers who have unauthorized access to the cloud and use data mining to extract information. Thus the privacy of data in the cloud has become a major concern in recent years. In this paper, we present an approach to prevent data mining based attacks on the cloud. Our system involves distributing user data among multiple cloud providers to make data mining a difficult job to the attackers. The key idea of our approach is to categorize user data, split data into chunks and provide these chunks to the proper cloud providers. In a nutshell our approach consists of categorization, fragmentation and distribution of data. The categorization of data is done according to mining sensitivity. Mining sensitivity in this context refers to the significance of information that can be leaked by mining. Categorization allows to identify sensitive data and to take proper initiatives to maintain privacy of such data. Fragmentation and distribution of data among providers reduce the

amount of data to a particular provider and thus minimize the risk associated with information leakage by any provider. This distribution is done according to the sensitivity of data and the reliability of cloud providers. The reliability of a cloud provider is defined in terms of its reputation. 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 sufficient number of chunks of data and thus prevents successful extraction of valuable information via mining. Even if an attacker manages to access required chunks, mining data from distributed sources remains a challenging job [28]. The main challenge in this case is to correlate the data seen at the various probes [30]. In addition to prevent data mining, the proposed system ensures greater availability of data and optimizes cost.

2. Basis and meaning of Data Mining

Organization and gathering can be used to sort out things by characterizing the common meaning among different things. The difficulty of data mining in centralized database, normally have the numerous following points: network traffic is measured less, mining productivity is low and the degree of spatial complication is high. The most typical classification data mining is classification methods based on distance, classification methods based on decision tree, Bayesian classification and so on.

Data mining techniques have been extensively used in various applications. However, the mistreat of these techniques may lead to the discovery of sensitive information. Researchers have recently made efforts at hiding sensitive association rules. However, undesired side effects, e.g., non-sensitive rules falsely hidden and spurious rules falsely generated may be formed in the rule hiding process. [5]

Privacy has become a significant issue in Data Mining. Many methods have been brought out to solve this problem. The basic aspect which we are concerned about in this paper is of association rule mining which preserves the confidentiality of each database. In order to find the association rule, each participant has to share their own data. Thus, a lot of privacy information may be put out or been illegally used. [6] Data mining can be defined as "the process that attempts to discover patterns in large data sets". The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use.

3. Applications

Exhausting this model, the clients or users can benefit different data mining services which the cloud providers promise to provide. As stated in the paper, the essential to progress this kind of model is the problem of sensitive data disclosure and expensive facilities which are come through when a general client-server model is used or when the

client's sensitive data is sent to the cloud while availing its services. By means of this model, these difficulties will not be challenged any more in the forthcoming. General client-server model is being used while communicating between a server and a client, both the client and the server need to share a common shared library so that they use the same constructs and formats to communicate.

This formed problematic since all the clients wanted to be complete conscious of this library if they are not. This twisted out to be costly, more time captivating and wasting of resources. Another aim why this model is being projected is that when a client wants to avail some of the services provided by the cloud, he/she needs to send their whole database to the cloud. This is done because for using the cloud's services the database of client is considered as input to the service routine. This leads to the sensitive information leak. Even if the database is encrypted in the cloud, then also the system administers and other officials can manage the decryption key. Hence, this technique also revolved out to be insecure and did not work.

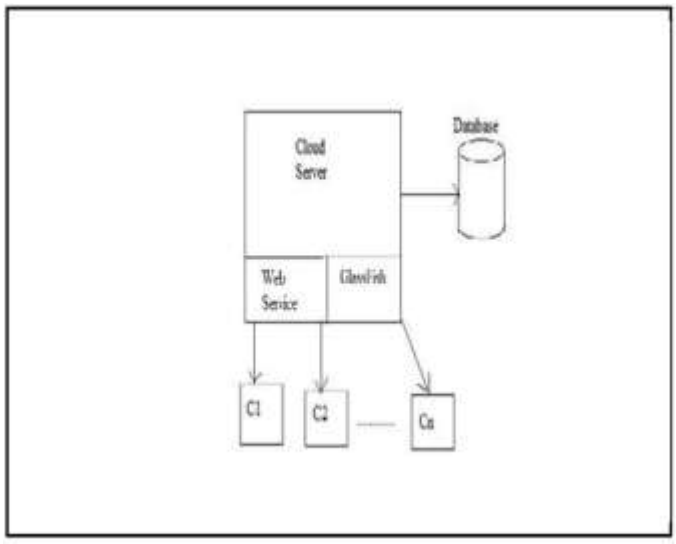

Figure 2: General Cloud Server Architecture

Web service is a method of communication over a network between two electronic devices. These were intended to solve three main problems such as Firewall Traversal, Complexity, and Interoperability.

The database represents the client's database which consists of the details about the transactions on the client's side. Software as a Service (SaaS) can be defined as a software distribution model in which applications are hosted by a vendor or service provider and made available to customers over a network, normally the Internet. It is becoming a gradually more widespread delivery model as underlying technologies that support Web services and service-oriented architecture (SOA) mature and new developmental approaches, such as Ajax, become popular. SaaS is closely related to the ASP (application service provider) and on demand computing software delivery models. Benefits of the SaaS model include: easier administration,

automatic updates and patch management, compatibility, easier collaboration, and global accessibility. The term "software as a service" (SaaS) is considered to be part of the classification of cloud computing, together with infrastructure as a service (IaaS) and platform as a service (PaaS). Most of SaaS solutions are based on a multi-tenant architecture. Using this model, a single version of the application, with a single configuration (hardware, network, operating system), can be used for all customers, i.e, tenants. To support scalability, the application is installed on multiple machines. Saas has an exception that some of its solutions do not use multi-tenancy, or use other mechanisms, such as virtualization, to cost-effectively manage a large number of customers in place of multitenancy.

GlassFish is a project started by Sun Microsystems which is an open-source application server for the Java EE platform. It is now sponsored by Oracle Corporation. The supported version of GlassFish is known as Oracle GlassFish Server. It is free software. It is the reference implementation of Java EE and also supports Enterprise JavaBeans, JPA, JavaServer Faces, JMS, RMI, JavaServer Pages, servlets, etc. This also allows developers to develop enterprise applications that are portable and scalable, and that combine with legacy technologies.

4. APPLICATION ARCHITECTURE

The application architecture of the proposed system is motivated by the Google file system. The Google File System is a scalable distributed file system for large distributed data-intensive applications [13].

When a client runs an application using files, the application can request for individual chunk by providing (client name, password, filename, sl no.) or for all chunks of a file by providing (client name, password, filename). In both the cases the password will have to be privileged enough to ask for the particular chunk(s).

If the privilege level of the password is greater than or equal to the privilege level of the chunk(s), the Cloud Data Distributor uses the chunk index field in the client table to identify the corresponding chunk(s) in the chunk table. The chunk table provides the virtual id of the corresponding chunk(s). It also provides the cloud provider index which identifies the corresponding provider entry/entries in the cloud provider table. The entry/entries of the cloud provider table provide(s) information regarding the provider(s) storing the chunk(s). After identifying the cloud provider(s), the Cloud Data Distributor uses the virtual id(s) as the key to obtain the required chunk(s) from the corresponding provider(s). Then the chunk(s) is(are) passed to the application. Consider a scenario from Figure 3 where a chunk request to Cloud Data Distributor is made using the quadruple (Bob, x9pr, file1, 0). Bob is listed as a client on Client Table and the password x9pr is listed under Bob. The privacy level of the password x9pr is 1 and the privacy level of chunk 0 of file1 is also 1. As the

privacy level of the password and the chunk is equal, the password is privileged enough to ask for the chunk. Now, the chunk index of chunk 0 of file1 is listed as 0 at Client Table. So the Cloud Data Distributor checks the 0th entry of Chunk Table which reveals the virtual id of the chunk, 10986. It also provides the current provider index 6 which in turn reveals the identity of the cloud provider from the Provider Table. The sixth entry of Cloud Provider Table is Earth. So, a chunk request to cloud provider Earth is made using 10986 as key. Upon receiving the chunk from Earth, the Cloud Data Distributor provides the chunk to the seeker. Consider another scenario where a request is made using quadruple (Bob, aB1c, file1, 0). The password aB1c is listed under Bob and its privacy level is 0. As the privacy level of the requested chunk is 1, the password is not privileged enough to access the chunk. Hence its request is denied.

Santos et al. [21] proposed a trusted cloud computing platform (TCCP) for ensuring the confidentiality and integrity of computations that are outsourced to IaaS services. The combination of our proposed system and the TCCP ensures the privacy of cloud data in case of outsourced storage and processing.

The present cloud storage system is a defenseless one because data stay under a single cloud provider. This can guide to data loss in case of proceedings like network outage, the cloud provider going out of business, malware attack etc. The present system also gives a huge benefit to the attackers as they have fixed targets in the forms of cloud providers. If an attacker chooses to attack a specific client, then he can aim at a fixed cloud provider, attempt to have way in to the client's data and examine it. This eases the job of the attackers. As long as the complete data belonging to a client remain under a single cloud provider, both inside and outside attackers get the benefit of using data mining to a great extent. Inside attackers in this context refers to malicious workers at a cloud provider. Data mining models frequently necessitate large number of comments and single provider structural design is a great benefit suiting the case as all the samples remain under the provider. Thus single provider architecture is the biggest security threat regarding data mining on cloud.

5. Conclusion

In cloud computing, the data is successful to be kept in storage area delivered by the service providers. The service providers must have a suitable technique to protect their client's complex data, principally to preserve the data from unlawful access. An acquainted method of information privacy protection is to store the client's data in encrypted form. If the cloud system is responsible for both storage space and encryption/decryption of the data, the system administrators may concurrently obtain encoded data and the decryption keys. This will permit them to access the information of the client without any approval.

These clues to the hazard of sensitive information reveal and the technique involved of storage and encryption/decryption is expensive too. Therefore, to rise above these difficulties, a model (cloud architecture) is explained in this paper which accepts only those data which are essential in an encoded form, performs the service opted by the client and sends the result in the encoded format to be implicit by the specific client.

References

[1] Sunil Sanka, ChittaranjanHota, MuttukrishnanRajarajan, "Secure Data Access in Cloud Computing," in IMSAA '10, 2010, p. 1-6.

[2] S. M. Mahajan and A. K. Reshamwala, "Data Mining Ethics in Privacy Preservation - A Survey " in International Journal of Computer Theory and Engineering, Vol. 3, No. 4, August 2011.

[3] Manoj Gupta and R. C. Joshi, "Privacy Preserving Fuzzy Association Rules Hiding in Quantitative Data" in International Journal of Computer Theory and Engineering, Vol. 1, No. 4, October, 2009.

[4] Jing-Jang Hwang and Hung-Kai Chuang, YiChang Hsu and Chien-Hsing Wu, "A Business Model for Cloud Computing Based on a Separate Encryption and Decryption Service," in ICISA '11, 2011, p. 1-7.

[5] Yi-Hung Wu, Chia-Ming Chiang, and Arbee L.P. Chen, "Hiding Sensitive Association Rules with Limited Side Effects," in IEEE Transactions on Knowledge and Data engineering, Vol. 19, No. 1,pp. 29-42, January 2007.

[6] Tinghuai Ma, Sainan Wang, ZhongLiu, "Privacy Preserving Based on Association Rule Mining," in Advanced Computer Theory and Engineering (ICACTE), Vol. 1, pp. 637-640, August 2010.

[7] Jiawei Han, MichelineKambe. Data Mining, Concepts and Techniques , 2nd Ed. CA: Morgan Kaufmann Publishers, 2006,pp. 234-239.

[8] Jiawei Han, MichelineKambe. Data Mining, Concepts and Techniques , 2nd Ed. CA: Morgan Kaufmann Pub

[9] Building Data Mining Applications for CRM [Paperback] by Alex Berson (Author), StephenJ.Smith (Author), Berson (Author), KurtThearling (Author).

[10] Data-Driven Marketing: The 15 Metrics Everyone in... by Mark Jeffery.

[11] Cloud Computing with the Windows Azure Platform By Roger Jennings

[12]Moving To The Cloud: Developing Apps in the New World of Cloud Computing, ByDinkarSitaram, GeethaManjunath

[13] The Cloud Computing Handbook - Everything You Need to Know about Cloud Computing, By Todd Arias

[14]http://searchsqlserver.techtarget.com/definition/datamining

[15]http://www.ijcaonline.org/volume15/number7/pxc387 2623.pdf

[16] http://www.waset.org/journals/waset/v39/v39-72.pdf

[17]http://www.estard.com/data_mining_marketing/data_mining_campaign.asp

[18]http://dssresources.com/books/contents/berry97.html

[19]http://www.marketingprofs.com/articles/2010/3567/th e-nine-most-common-data-mining-techniques-usedin-predictive-analytics.

[20] http://www.thearling.com/