A SURVEY ON DATA SECURITY IN CLOUD COMPUTING: ISSUES AND MITIGATION TECHNIQUES

Satarupa Biswas[1], Abhishek Majumder[2]

[1]*Department of Computer Science and Engineering, Tripura University*
[2]*Department of Computer Science and Engineering, Tripura University*

Abstract

Cloud is considered as the future of information technology. Cloud computing refers to internet based computing where virtual servers provide software, hardware, infrastructure, devices and platform. Users can make them utilized on a pay-as-you-use basis. Many organizations are stringent about the adoption of cloud computing due to their concern regarding the security of the stored data. Therefore, issues related to the security of data in the cloud have become very vital. In this paper, different data security issues have been discussed. Moreover, some already existing schemes for countering the problem of data security have also been discussed. Finally, a comparative study has been carried out among the schemes with respect to data security issues.

Keywords- *Cloud computing, Data security, Servers, Database, Multi-Clouds Database Model (MCDB)*

----------***----------

1. INTRODUCTION

Now-a-days, Cloud Computing is the most popular up-coming concept in Information Technology (IT). The name Cloud Computing gets derived from the cloud like structure that was used to represent internet. It is the style of computing where massively scaled IT related capabilities are provided as a service across the internet to multiple external customers and are billed by consumption. It is a recent technology, where users need not to have their own hardware, software, storage space etc. Everything will be provided by the cloud itself. Google, Microsoft, Yahoo, IBM and Amazon have started providing cloud computing services. Amazon is the pioneer in this field. Cloud Computing provides the facility to access shared resources and common infrastructure offering services on demand over the network to perform operations. But the location from where the users accessing the requisite data are not known. With Cloud Computing, users can access their databases from anywhere in the world only when they remain connected to the internet.

Today's world relies on cloud computing to store their public as well as personal data. That data may be required by them or others at any instant of time. As a result, data security in cloud computing has drawn lots of attention from the research community.

The rest of the paper is organized as follows. Section 2 gives an overview on delivery and deployment models of cloud computing. Section 3 discusses the issues related to data security in cloud computing. Some of the techniques for mitigating the problem of data security in cloud computing have been described in section 4. In section 5, a comparative study of data security mitigation algorithms that are discussed in section 4, have been carried out considering the data security issues. Finally, section 6 presents the conclusion.

2. OVERVIEW

Cloud Computing is an idiom that involves hosted services over the internet. Cloud Computing provides the facility to access shared resources and common infrastructure, offering services on demand over the network to carry out operations that meets changing business needs. Cloud Computing allows consumers and businesses to use applications without installation and can access their personal files at any computer with internet access. The location of physical resources and devices being accessed are usually not known to the end user. It also provides facilities for users to develop, deploy and manage their applications in the cloud, which requires virtualization of resources that maintains it. There are a small number of differences between traditional hosting and cloud computing. Cloud Computing is sold on demand. All the services in the cloud are managed by the provider. Users can determine the amount of services they take and can log on to the network from any computer in the world.

2.1 Delivery Models

The architecture of Cloud computing can be categorized according to the three types of delivery models [13][15]:

- Infrastructure as a Service (IAAS): Consumers are allocated computing resources in order to run virtual machines that consist of operating systems and applications that are provided as an on-demand service. The best example of IAAS is Amazon.com's Elastic Compute Cloud (EC2) service. The requirements of security beyond the basic infrastructure are carried out mainly by the cloud consumer.

- Platform as a service (PAAS): Cloud consumers are allowed to write applications that run on the service provider's environment. It is a model of service delivery where the computing platform is provided as an on-demand service upon which applications can be developed and deployed. Google Apps engine is an example of PAAS. The requirements of security are split between the cloud provider and the cloud consumer.
- Software as a service (SAAS): Cloud consumers are provided with various software applications that run over the internet. Google Docs programs are an example of SAAS. The requirements of security are carried out mainly by the cloud provider.

2.2 Deployment Models

Deployment models broadly characterize the management and disposition of computational resources for delivery of services to consumers, as well as the differentiation between classes of consumers. Different deployment models are [11] [15] [16][19]:

- Public Cloud: A public cloud is one in which the infrastructure and computational resources that it comprises are made available to the general public over the Internet. It is owned and operated by a cloud provider delivering cloud services to consumers and, by definition, is external to the consumers' organizations.
- Private Cloud: A private cloud is one in which the computing environment is operated exclusively for a specific organization. It may be managed by the organization or by a third party, and may be hosted within the organization's data center or outside of it.
- Community Cloud: A community cloud falls between public and private clouds. It is somewhat similar to a private cloud, but the infrastructure and computational resources are exclusive to two or more organizations that have common privacy, security, and regulatory considerations, rather than a single organization.
- Hybrid Cloud: Hybrid cloud is the most complex model among all the deployment models. They involve a composition of two or more clouds (private, community, or public). Each cloud remains a unique entity, but is bound to the others through consistent or proprietary technology that enables application and data portability among them.

3. DATA SECURITY

Data seeking is done immensely by many cloud consumers, which can give rise to serious security concerns in cloud environment. In this section, different issues related to data security in cloud computing have been described.

There is a critical need to securely store, manage, share and scrutinize enormous amounts of data. It is very important that the cloud should be secured enough to maintain the security of data. Exact physical localization of user data in virtual cloud atmosphere is among some of the prime challenges in cloud computing. The major security challenge with clouds is that the owner of the data may not have complete knowledge of where their data are stored. Data security involves encrypting the data as well as ensuring that suitable policies are imposed for sharing those data. There are numerous security issues for cloud computing. Some of the major data security issues are [13]:

- Data Integrity: It is very essential to maintain the integrity of data. The stored data in the cloud storage may suffer from enormous damage occurring during the transition operations from or to the cloud storage provider. The risk of attacks from both inside and outside the cloud provider exists and should be considered.
- Data Intrusion: Data Intrusion is another security risk that may occur with a cloud provider. If any intruder can gain access to the account password, then he/she will be able to do any kind of unwanted changes to the account's private documents. Undesirable alteration of user data may commence due to intrusion.
- Service availability: Service availability is another major concern in cloud services. It is mentioned in some cloud providers licensing agreement, that the service may be unavailable anytime due to some unforeseen reason. If all the valuable business documents are stored on the cloud and the cloud suddenly goes down, will it be coming back up with all of our important documents intact? It is also important to know whether the company with whom user is storing his vital information is financially steady and will not suddenly vanish taking all of the valuable information with it.
- Confidentiality: Confidentiality of data is another security issue associated with cloud computing. The data should be kept secured and should not be exposed to anyone at any cost. The users do not want their confidential data to be disclosed to any service provider. But it is not always possible to encrypt the data before storing it in cloud.
- Non- Repudiation: Non-repudiation is a major concern for data security. It guarantees the transmission of message between parties and gives the assurance that someone cannot deny something. Non-repudiation is often used for signatures, digital contracts, and email messages. It ensures that a party cannot deny the genuineness of their signature on a document or the sending of a message that they originated.

4. MITIGATION ALGORITHMS

In this section some of the data security algorithms have been discussed. Users are benefited by the cloud service providers, which may give rise to new security risks. Users storing their data in cloud are not aware of the physical location of their data.

Attacks on the data storage can directly affect the security of the user's valuable data including application data or sensitive data. Many cloud service providers provide storage as a service. They take the data from the users and store them on large data centers. The data stored in the cloud may get lost or modified due to some security breaches that may be caused by many factors. Several algorithms for ensuring data security have been proposed.

Pradeep Kumar et al. [1] proposed a scheme for security in cloud computing using Hidden Markov model (HMM) [17] and Clustering. This scheme performs intrusion detection based on the probability of the behavior. Clustering is used for providing only those data that are required by the user. Cluster provides a compact view of data in cloud environment. It is used for fraud detection and is utilized to reduce the data seeking time [18]. In [1], HMM model monitors the user behavior continuously and informs the administrator with the help of filter network. Firewall gateway is used here to block the SSH service, which is used for gaining access to the server from anywhere in the world. It simply drops the packets that come from any IP address which is not inside its knowledge. It is the responsibility of the system administrator to keep the data safe. The system administrator immediately takes the action if any malicious activity related to data mining is detected by the filter network after getting input from the HMM. In this scheme, the cloud computing environment has a plug-in attached with it. It can help the proposed model to backup as well as recover data in case of emergency. It can be easily connected with any other cloud environment if needed. As a result, the load on the servers of a cloud could be minimized. So, an intruder can be easily caught even if he has the stolen ID and/or password. This environment has classified the data into two categories. One category includes normal data seeking from the database. On the other hand, Second category involves seeking sensitive data that must be kept protected from unauthenticated access.

Palivela Hemant et al. [2] have introduced a new prototype system where a central server will save all the information in its router table that can be useful for back tracking the server or user. The router table contains cloud id, user id, the actual server id to which the user is connecting, server name, total time of synchronization, packet size, lease time, source ip and destination ip. It also contains the packets per second transfer rate which is the actual amount of data flow. There will be application level firewalls which will not only check the undesirable web sites based on their ip addresses but they will also keep track if the packets are malicious. They can record the activities of the user. The user side contains personal firewall and the connectivity between user and the central server will be encrypted using SSL encryption standards. When any user wants to connect to a particular server then his/her information gets stored in the table. In case, the user is unable to connect to the server, the server can be easily backtracked from the central server's routing table. Many of the security requirements are fulfilled in this model due to double encryption.

Shuai Han et al. [5] proposed a third party auditor scheme in cloud computing for ensuring data storage security. Third party auditor (TPA) provides trustful authentication for the user who stores their data in the cloud. TPA is powerful than the cloud consumers and makes every data access be in control. Users cannot completely rely whether their data is safe on the cloud providers. TPA can review and interpret the data stored in the cloud on behalf of the users upon request. It can provide a log report to the users. The authors have proposed a new architecture for cloud storage, where the third party auditor and the cloud service provider have been combined together. The traditional network architecture [7] consists of three entities which are users, Cloud service provider and TPA. Users have large data files to be stored in the cloud. They are active participants and can be individual consumers or organizations. Cloud service provider has sufficient storage space as well as computation resources for maintaining the data stored by the users. In this scheme RSA has been used to encrypt the data flow between servers in the advance cloud service provider. It uses Bilinear Diffie-Hellman algorithm for exchanging keys. Users and cloud service provider can communicate with each other using a message header without a third party auditor. Each of the cloud storage servers can add, identify and update the message header for users whose copy will be sent to the trustful organization server for the first time. A pair of keys is allocated to each user for accessing the cloud. In this scheme, users add a message header before sending it to the cloud. The data packets are encrypted with the allocated keys using RSA algorithm. Trustful organization servers which are maintained by trustful organization's that perform as a watchdog for every access keys in cloud service provider, comprises of few number of servers. The users and cloud service provider cannot get any authentication information from trustful organization without a certain module. The Trustful organization server is responsible for maintaining all the keys which are stored in cloud storage servers.

According to Mohammed A. Alzain et al. [4], shifting from single cloud to multi-cloud is very important for ensuring the security of user's data. Authors suggested that, there are three main security factors of data (data integrity, data intrusion and service availability) that needs to be considered as the major concern for cloud computing. They have proposed a new model called Multi-clouds Database Model (MCDB). A technique named Shamir's secret sharing algorithm [3], which is based on polynomial interpolation has been incorporated in the scheme. According to the algorithm [3], if a data D is shared into n pieces, in such a way that D is easily reconstruct able from k pieces, but even complete knowledge of k-1 pieces reveals absolutely no information about D. The authors have suggested that Cloud Computing should not end with a single cloud. In their work, they have compared Amazon cloud service which is single cloud with their proposed multi-clouds model. This model guarantees the security and privacy of data in multi-clouds using multi shares technique instead of single cloud. The data is replicated among several clouds by using secret sharing

approach. The operations between the clients and the cloud service providers are controlled by Database Management System (DBMS). Data is being stored by cloud service providers after being divided by MCDB. Division of the data depends upon the number of cloud service providers.

Mohammed A. Alzain et al. [9] proposed a Multi-clouds Database Model (MCDB) which uses multi-clouds instead of single cloud service provider such as in Amazon cloud service. In addition, it employs Shamir's secret sharing approach to ensure security of the stored data in the cloud. Furthermore, it adopts a triple modular redundancy (TMR) technique which is a type of passive hardware redundancy. In TMR technique, three identical modules execute the same task in parallel. If one of the three models was faulty, the other two models will mask and hide the result of the faulty module. This technique is used with sequential method to improve the reliability of the system and multi shares technique to improve the security of the system. The sequential voting method decreases the number of execution cycles. When the data is divided by the cloud manager, then it will be sent and stored directly into the clouds. There is no need for storing a copy of the user data in the cloud manager. It is the function of cloud manager component to generate and compute the polynomial functions. Thereafter on the basis of majority voting of the output results that has come from clouds the faulty cloud inside the super cloud provider is detected. All the shares from different clouds will go through voter inside the cloud manager. The result of the voting does not get affected by the execution of the third cloud's result, if two results out of the three clouds were same.

5. COMPARISON

This section presents a comparison among the data security algorithms, discussed in the previous section, with respect to the issues discussed in section III.

Table 1: Various Schemes and their support of features

Addressed Security Risks	[2]	[9]	[4]	[1]	[5]
Data Integrity	√	√	√	√	X
Data Intrusion	X	√	√	√	√
Service Availability	√	√	√	√	X
Data Confidentiality	√	√	√	√	√
Non-Repudiation	√	X	X	√	√

Table 1 is depicting the various security risks that have been resolved ([2],[9],[4],[1],[5]) by tick sign in the corresponding field of the table. In [2], hashing is done in SSL technology which helps to maintain the Integrity of data. In this scheme, confidentiality is achieved by using dual SSL technology. Another security risk called data intrusion that may occur with a cloud service provider (CSP) is kept unsolved in this scheme. Authentication through conventional SSL can be weak and subject to man-in-the-middle attacks. In this scheme, service availability issue is solved using the central server that acts as a backup server. So, if the server goes down, the failure can be avoided. Non-repudiation is satisfied using the central server, which stores all the necessary information about a user in its routing table.

In [4] and [9], few major security issues such as, data integrity, data confidentiality, service availability and data intrusion have been solved. In [9], TMR technique helps to maintain integrity by detecting the availability of cloud and it can determine the faulty cloud. Data integrity is maintained in [4], because [3] ensures that, if a data D is divided into n pieces. Data D can only be reconstructed when a sufficient number of shares are combined together. That is individual shares are of no use on their own. In [4] and [9], data confidentiality is maintained by storing the data in multiple cloud service providers and clouds respectively by using Shamir's secret sharing approach [3]. In [9], TMR techniques are used to resolve the Data Intrusion issues. In this scheme, it is very easy to detect the faulty cloud and can explore where the intrusion has taken place. Intrusion can be prevented in [4], because if the hacker hacked the password from one cloud service provider, they still have to hack the third cloud service provider for its password. It is very hard for a hacker to retrieve the password from all the cloud service providers. Service availability is guaranteed in [4] and [9], as the data is distributed in different cloud service providers and clouds respectively. As a result, the risk of data loss gets minimized. The problem of non- repudiation have not been taken care of in the schemes [4] [9].

In [5], data intrusion problem have been solved, as, only the authenticated users can have access to the information. The users are given permission using the authentication modules which are given by the trustful organizations. Any intruder cannot have access to the valuable data without proper permission. Confidentiality is achieved as the data file is encrypted with user secret key. After encryption, the data file is sent for storing in the cloud service provider. Service availability and data integrity problems are not considered in this scheme. If the CSP goes down, then the data cannot be recovered. The risk of attacks exists in the Cloud storage provider. The data stored in the cloud may suffer from any damage occurring during transition from or to the cloud storage provider. But, no such measures have been taken in this scheme to ensure the integrity of data. Non-repudiation is not possible here, because when a user sends a request, then the user's information is added in cloud storage server's user list.

In [1], data integrity is maintained by using HMM model which monitors the user behavior continuously and informs the system administrator. As a result, the data is safe in cloud environment. HMM model can be used to detect intrusion based on the probability of behavior [17]. Firewall gateway is also used in this scheme [1] to prevent data intrusion. Firewall gateway can block any anonymous port with a live IP address which is not inside its knowledge. As a result, intrusion is not possible in this

scheme. Data can be kept confidential using this scheme, as every traffic update is sent to the administrator via email notification. Service availability can be solved with the help of plug-in, which helps the environment to connect with another network. Non- repudiation is not possible as the firewall gateway contains the list of authenticated IP addresses.

6. CONCLUSIONS

Data security is a very critical issue in cloud computing. In cloud data storage system, users store their data in the cloud cannot possess the data locally. Users are not aware of the physical location of their data. It is not clear how safe their data is and ownership of data is also unclear when these services are used. Cloud computing companies say that the data stored are completely safe. But, it is too early to comment on the reliability issues claimed by them. The stored data may suffer from damage that occurs during data transition operations from or to the cloud provider. Data are not always safe when they are stored inside cloud providers. For addressing these issues several algorithms have already been proposed and there is a huge scope for work in the area of data security in cloud computing.

REFERENCES

[1] Pradeep Kumar, Nitin, Vivek Sehgal, Kinjal Shah, Shiv Shankar Prasad Shukla and Durg Singh Chauhan, "A Novel approach for Security in cloud computing using Hidden Markov model and Clustering", *Proc. of Information & Communication Technologies (WICT)*, pp. 810-815, 2011.

[2] Palivela Hemant, Nitin.P.Chawande, Avinash Sonule, Hemant Wani. "Development of servers in cloud computing to solve issues related to Security and Backup", *Proc. of IEEE international Conference on Cloud Computing & Intelligence Systems (CCIS)*, pp. 158-163, 2011.

[3] A. Shamir, "How to share a secret", Communications of the ACM, Vol. 22, Issue 11, pp. 612-613, 1979.

[4] Mohammed A. Alzain, Ben Soh and Eric Pardede, "MCDB: Using Multi Clouds to ensure Security in Cloud Computing", Proc. of the 2011 IEEE 9th International Conference on Dependable, Autonomic & Secure Computing (DASC), pp. 784-791, 2011.

[5] Ian Foster, Yong Zhao, Ioan Raicu and Shiyong Lu, "Cloud Computing and Grid computing 360-Degree Compared", Grid Computing Environments Workshop, 2008, GCE'08 , pp. 1-10,2008.

[6] Shuai Han , Jianchuan Xing, "Ensuring data storage security through a novel third party auditor scheme in cloud computing", IEEE international Conference on Cloud Computing & Intelligence Systems(CCIS), pp. 264-268, 2011.

[7] Qian Wang, Cong Wang, Jin Li, Kui Ren, Wenjing Lou, "Enabling Public Verifiability and Data Dynamics for Storage Security in Cloud Computing", Proceedings of the 14th European conference on Research in computer security, pp. 355-370,2009.

[8] Abhishek Mohta and Lalit Kumar Awasthi. "Cloud Data Security using Third Party Auditor", International Journal of Scientific & Engineering Research (IJSER), Vol. 3, Issue 6, 2012.

[9] Mohammed A. Alzain, Ben Soh and Eric Pardede. "A new approach using redundancy technique to improve security in cloud computing". Cyber Security, Proc. of International Conference on Cyber Security, Cyber Warfare & Digital Forensic(Cybersec), pp. 230-235, 2012.

[10] P. Mell and T. Grance. "The NIST definition of cloud computing". National Institute of Standards and Technology 2009.

[11] Zaigham Mahmood,"Data Location and Security Issues in Cloud Computing", Proc. of International Conference on Emerging Intelligent Data and Web Technologies (EIDWT), pp. 49 – 54, 2011.

[12] Pengfei You, Yuxing Peng, Weidong Liu, Shoufu Xue, "Security Issues and Solutions in Cloud Computing", Proc. Of 32nd International Conference on Distributed Computing Systems Workshops, pp. 573-577, 2012.

[13] Mohammed A. AlZain, Eric Pardede, Ben Soh, James A. Thom, "Cloud Computing Security: From Single to Multi-Clouds", Proc of the 45th Hawaii International Conference on System Sciences (HICSS), IEEE, pp. 5490-5499, 2012.

[14] Veerraju Gampala, Srilakshmi Inuganti, Satish Muppidi, "Data Security in Cloud Computing with Elliptic Curve Cryptography", Proc. of International Journal of Soft Computing and Engineering, Vol 2, Issue 3, pp. 138-141, 2012.

[15] Pengfei You, Yuxing Peng, Weidong Liu, Shoufu Xue , "Security Issues and Solutions in Cloud Computing", in Proc. of 32nd International Conference on Distributed Computing Systems Workshops(ICDCSW), pp. 573-577, 2012.

[16] T.M Bhraguram , M.S Sumesh, "Cyber Security Information Exchange Based on Data Asset De-coupling factor in Cloud Computing", IEEE Recent Advances in Intelligent Computational Systems (RAICS), pp. 089 – 095, 2011.

[17] Abhinav Shrivastava, Amlan Kundu, S Surat, A.K. Majumdar "Credit Card Fraud Detection Using Hidden Markov Model", IEEE Transactions on Dependable And Secure Computing, vol. 5, issue 1, pp. 37-48, 2008.

[18] Jeffrey W. Seifert, "Data Mining: An Overview", CRS Report for Congress, pp.1-16, 2004.

[19] Pavithra S., Badi Alekhya, "Implementing Efficient Monitoring And Data Dynamics For Data Storage Security in Cloud Computing", IRACST - International Journal of Computer Science and Information Technology & Security (IJCSITS), pp. 139-143, Vol. 2, No. 1, 2012.