Indian Journal of Science and Technology, *Vol 8(24), DOI: 10.17485/ijst/2015/v8i24/84229, September 2015*

ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

A Study of Data Storage Security Issues in Cloud Computing

M. B. Jayalekshmi[1]* and S. H. Krishnaveni[2]

[1]Department of Computer Science and Engineering, Noorul Islam University, Kumaracoil,Kanyakumari- 629180, Tamil Nadu, India; mbjaya2k5@gmail.com
[2]Department of Information Technology, Noorul Islam University, Kumaracoil, Kanyakumari- 629180, Tamil Nadu, India; shkrishnaveni@gmail.com

Abstract

Cloud Computing is defined as an environment in which users can share their resources with others in pay per use model. The resources are stored centrally and can access from anywhere. Despite these advantages, there still exist significant issues that need to be considered before shifting into cloud. Security stands as major obstacle in cloud computing. This paper gives an overview of the security issues on data storage along with its possible solutions. It also gives a brief description about the encryption techniques and auditing mechanisms.

Keywords: Confidentiality, Encryption, Integrity, Privacy, Security

1. Introduction

Cloud computing nowadays is an emergent IT technology which has gained limelight in research. Cloud computing is the combination of many pre-existing technologies that have matured at different rates and in different contexts. The goal of cloud computing is to allow users to take benefit from all these technologies. Many organizations are moving into cloud because it allows the users to store their data on clouds and can access at anytime from anywhere. Data breaching is possible in cloud environment, since data from various users and business organizations lie together in cloud. And also by sending the data to the cloud, the data owners transfer the control of their data to a third person that may raise security problems. Sometimes the Cloud Service Provider(CSP) itself will use/corrupt the data illegally.

Security and privacy stands as major obstacle on cloud computing i.e. preserving confidentiality, integrity and availability of data. As simple solution encrypt the data before uploading it onto the cloud. This approach ensures that the data are not visible to external users and cloud administrators but has the limitation that plain text based searching algorithm are not applicable. In this paper we discuss the security flaws in data storage and also the mechanisms to overcome it. The rest of the paper is organized as follows: Section 2 gives an overview of cloud computing. Section 3 describes the threats in cloud computing. The security and privacy issues of data storage and an introduction to encryption and auditing were explained in Section 4 and the final conclusions are summarized in Section 5.

2. Overview

The National Institute of Standard and Technology's (NIST) defined cloud computing as a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction. The NIST has listed five main characteristics of cloud computing as:

- **On-demand self-service:** Resources are available to users based on their demand.

** Author for correspondence*

- **Broad network acess:** The services are rendered over the network and the users can access it if having an internet connection.
- **Resource pooling**: Resources from vendors are pooled to serve multiple users.
- **Rapid elasticity:** Users can access the resources whenever needed and also they can release the resources when they no longer required.
- **Measured service:** Users have to pay only for the time they are using the resources.

The delivery models in cloud shown in Figure 1. They are:

- **Infrastructure as a Service (IaaS):** The IaaS model offers the infrastructure to run the applications.
- **Platform as a Service (PaaS):** The PaaS model enables the application developer with a development environment and also offer the services provided by vendor.
- **Software as a Service (SaaS):** In SaaS model, the users can use the software for rent instead of purchasing it.

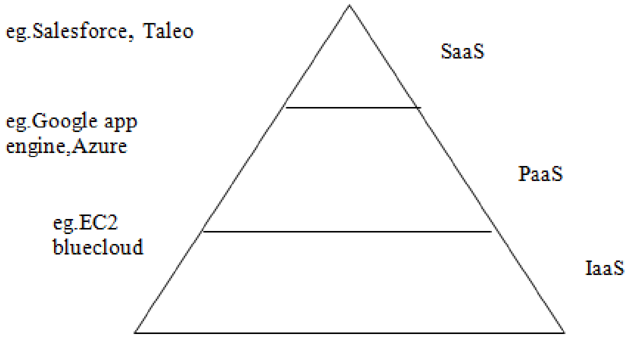

Figure 1. Delivery models.

The deployment models in cloud are:

- **Public Cloud:** The cloud is said to be public cloud, if its services are rendered for open use by the general public. It may be owned, managed and operated by a business, academic, or government organization, or some combination of them. Amazon, Google are examples of public cloud.
- **Private Cloud:** The cloud is said to be private cloud, if it is owned and managed solely by an organization and its services are rendered to the users within the organization.
- **Community Cloud:** A community cloud is an infrastructure shared by several organizations which supports a specific community.
- **Hybrid Cloud:** A hybrid cloud is a combination of public and private clouds.

3. Threats in Cloud Computing

There are certain aspects associated with Cloud Computing as a result of which many organizations are still not confident about moving into the cloud. The Computer Security Alliance Group has listed the threats that may occur in cloud computing. They are:

- Abuse of cloud computing.
- Insecure Interfaces and API's.
- Malicious Insiders.
- Shared Technology Issues.
- Data Loss and Leakage.
- Account or Service Hijacking.
- Unknown Risk Profile.
- Hardware Failure.
- Natural Disasters.
- Closure of Cloud Service.
- Cloud-related Malware.
- Inadequate Infrastructure Design and Planning.

Among these data loss and leakage was ranked as the second most common threat. Data loss and leakage occurs due to lack of security and privacy in both storage and transmission. To reduce this risk, the data security aspects taken into account are:

- **Data-in-transit:** Data-in-transit refers to the data during transmission either from data owner to cloud provider or from cloud provider to owner.
- **Data-at-rest:** Data-at-rest refers to the data in the storage.
- **Data lineage:** Data lineage specifies what happened to data from its source through distinct applications and its use for auditors. Data lineage is difficult for public clouds.
- **Data provenance:** Data provenance is not just proving the integrity of data, but the more specific history of the data i.e., who created, modified and deleted the data in the cloud.
- **Data remanence:** Data remanence refers to the data left behind after deletion[1].

This paper highlights the issues related to data storage. Data Storage refers to storing the data on a remote sever hosted by the CSP. The benefits of data storage in cloud are:

- Provides unlimited storage space for storing user's data.
- User can access the data at anytime from anywhere using an internet connection in more than one machine.
- No need to buy the storage device for storing the data.

The main constraint in data storage was absence of security and privacy which arises due to loss of control over the data. The basic requirements for secure data storage are:

- The data on the cloud must be confidential and CSP should not be able to compromise it at any cost.
- Data access must be given to the intended user only.
- The data owner must have full control over the authorization of data[2].

4. Security and Privacy Issues in Data Storage

Cloud Computing allows the users to store their data on the storage location maintained by a third party. Once the data is uploaded into the cloud the user loses its control over the data and the data can be tampered by the attackers. The attacker may be an internal(CSP) or external. Unauthorized access is also a common practice due to weak access control. The protection of information arises the following challenges:

- Access control: Are there appropriate controls over access of information when stored in the cloud?
- Structured versus unstructured: How is the data are stored? Whether it supports data access in a very fast manner?
- Integrity/availability/confidentiality: How are data integrity, availability and confidentiality maintained in the cloud?
- Encryption: Several laws and regulations require that certain types of information should be stored only when encrypted. Is this requirement supported by the CSP?

The security and privacy issues related to data storage are confidentiality, integrity and availability.

4.1 Confidentiality

The major dispute in cloud computing is confidentiality. Data confidentiality means accessing the data only by authorized users and is strongly related to authentication. In another way confidentiality means keeping users data secret in the cloud systems. As we are storing the data on a remote server and transferring the control over the data to the provider here arises the questions such as:

- Will the sensitive data stored on the cloud is confidential?
- Will the cloud provider itself be honest?

For ensuring confidentiality, cryptographic encryption algorithms and strong authentication mechanisms can be used. Encryption is the process of converting the data into a form called cipher text that can be understood only by the authorized users. Encryption is an efficient technique for protecting the data but have the obstacle that data will be lost once the encryption key is stealed[3]. The major potential concern is:

- How the data in the cloud is be protected?
- If encryption is used what will be its key strength?

It all depends on the CSP. CSP itself will encrypt the user data before storing and the keys will be disclosed only to the authorized persons. But some CSPs allow the users to encrypt their data before uploading into the cloud. The encrypted data is usually stored in the server and the keys are revealed only to the authorized users[4]. Different cryptographic algorithms are available for encryption. In symmetric encryption involves the use of private key is used for both encryption and decryption as shown in Figure 2. In symmetric algorithms the data is encrypted by using a private or secret key and the same key is used for decryption also. Symmetric algorithms include DES, AES and Blowfish etc. In[5] DES has been a popular symmetric key encryption, introduced in 1976 and is used in many commercial and financial applications. DES is easier to implement in both hardware and software but is slower and has poor performance. DES was replaced by AES encryption which is fast and flexible and was used to protect information in smart cards and online transactions. The key size of 256 bits is more secure but sometimes it is too complex. Blowfish introduced in 1993 is one of the most common public domain encryption algorithms. Blowfish is a fat and simple encryption algorithm.

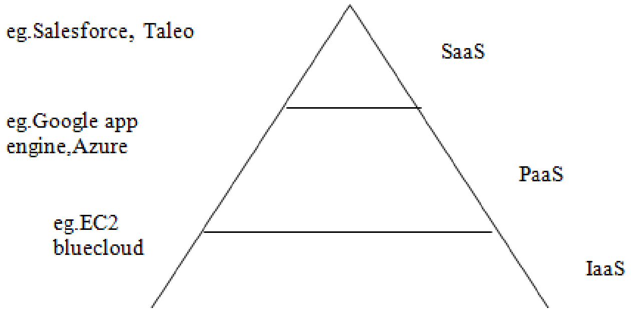

Figure 2. Symmetric encryption.

In general symmetric algorithms are simpler and faster but not efficient that both sender and receiver share the same secret or private key.

Asymmetric encryption algorithms also called public key encryption involves the use of public key and private key. In asymmetric encryption algorithms the sender encrypts the data using public key of the receiver and the receiver will decrypt it using his private key. The most popular asymmetric encryption algorithm is RSA encryption which is developed in 1978. It provides increased security as the private keys do not need to be revealed to anyone. Another advantage is it provides mechanisms for digital signature. Digital signatures along with RSA encryption ensure security of data in cloud. A digital signature is a mathematical scheme for proving the authenticity of data.

Predicate encryption is also a kind of asymmetric encryption which allows decrypt selected data instead of decrypting all of it. Identity Based Encryption(IBE) is a public key encryption which uses the unique information about the identity of the user as public key and guarantees authenticity. The major advantage of asymmetric encryption is it provides more security. The disadvantage is its speed i.e., symmetric algorithms are faster than asymmetric algorithms. Figure 3 depicts the asymmetric encryption technique.

The above encryption techniques have the limitation that for searching the data from the file, the entire file has to be decrypted. It is a time consuming process and thus searchable encryption was introduced. Searchable encryption allows build an index for the file containing the keywords and is encrypted and stored along with the file, so that while searching the data only the keywords are decrypted rather than the entire file and search is made on it[6].

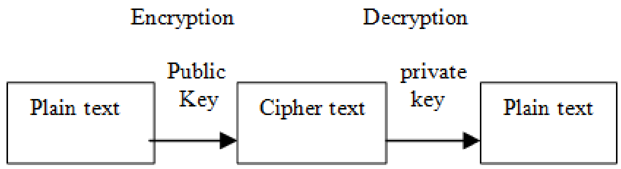

Figure 3. Asymmetric encryption.

An efficient encryption is homomorphism encryption which allows the CSP to carryout operations on encrypted file rather than decrypting it, which provides the same result[7]. The key used for encryption is kept secret by the user and not revealed to the CSP, so it is more secure.

All these encryption algorithms will improve security of data but maintaining the encryption key as secret is a difficult tasks for the CSP as more users dumping their data. As the key is with the CSP sometimes it is possible to hack the data.

4.2 Integrity

Another serious problem faced by cloud computing is integrity. Integrity of data means to make sure that the data has not been changed by an unauthorized person or in an unauthorized way. It is a method for ensuring that the data is real, accurate and safeguarded from unauthorized users. As cloud computing supports resource sharing, there is a possibility of data being corrupted by unauthorized users. Digital Signatures can be used for preserving the integrity of data. The simple way for providing integrity is using Message Authentication Code(MAC). Message Authentication Code is a cryptographic checksum calculated using hash functions and is send along with the data for checking the integrity. Auditing mechanisms can also be used for preserving integrity. In private auditing the integrity of data is checked by the data owner using algorithms. Public auditing means assigning a Trusted Third Party (TPA) by the data owner to check the integrity of the data. The TPA cannot access the data but can verify whether the data is modified or not and will report to the owner.

Remote Data Auditing refers to a group of protocols for verifying the correctness of the data over the cloud managed by CSP without accessing the data. As shown in Figure 4 Remote Data Auditing follows response-challenge process which involves the following steps:

- The data owner processes the file and generate meta data and handover it to the TPA.
- The TPA generates a challenge and transmits to CSP for checking the data correctness.
- On receiving the challenge the CSP calculates the response and send it to TPA.
- After receiving the response verification is done by TPA to check whether the data is stored correctly by the provider[8].

Provable Data possession is also a remote auditing mechanism. In all PDA mechanisms the data owner or TPA will check the integrity of data. However TPA is not able to check the integrity independently when the data owner fails to send the metadata for verification. The TPA does not have the permission to take countermeasures without informing the owner. To overcome this issue proxy PDP was defined in which remote data auditing task was assigned to a proxy on warrant[9].

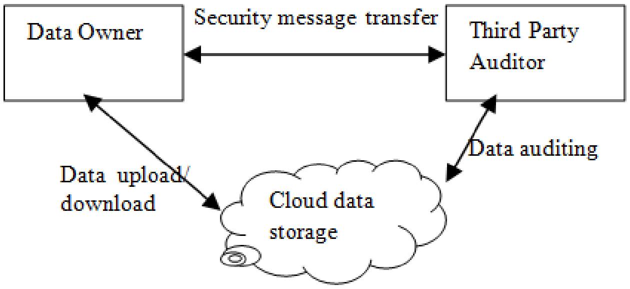

Figure 4. Remote auditing mechanism.

4.3 Availability

Availability refers to being available and accessible to authorized users on demand. The aim of availability in cloud computing systems is to ensure that its users can use them at any place and at any time.

5. Conclusion

Cloud computing enables users to store their data in remote storage location. But data security is the major threat in cloud computing. Due to this many organizations are not willing to move into cloud environment. To overcome this, confidentiality, integrity, availability should be encapsulated in a CSP's Service-Level Agreement (SLA) to its customers. Otherwise ensure that any sensitive information is not put into a public cloud and if any it is to be stored in encrypted form. Effective auditing mechanisms also can be used for providing data integrity.

6. References

1. Mather T, Kumaraswamy S, Latif S. Cloud security and privacy, an enterprise perspective on risks and compliance. O'reilly. First ed. 2009.
2. Sookhak M, Talebian H, Ahmeda E, Gania A, Khurram-Khan M. A review on remote data auditing in single cloud server: Taxonomy and open issues. J Netw Comput Appl. Elsevier. 2014; 43:121–41.
3. Sen J. Security and privacy issues in cloud Computing. Conference
4. Rahmani H, Sundararajan E, Md. Ali Z, Mohd Zin A. Encryption as a Service (EaaS) as a solution for cryptography in cloud. International Conference on Electrical Engineering and Informatics; Procedia Technology; 2013. p. 1202–10.
5. Mitali VK, Sharma A. A survey on various cryptography techniques. International Journal of Emerging Trends and Technology in Computer Science. July-Aug 2014; 3(4):6. ISSN 2278-6856.
6. Hana F, Qin J, Zhaob H, Hu J. A general transformation from KP-ABE to searchable encryption. Journal Future Generation Computer Systems. 2014; 30:107–15.
7. Tebaa M, El Hajii S. Secure cloud computing through homomorphic encryption. International Journal of Advancements in Computing Technology. 2013 Dec; 5(16).
8. Rong C, Nguyen ST, Jaatun MG. Beyond lightning: A survey on security challenges in cloud computing. Journal Computers and Electrical Engineering. 2013; 39(1):47–54.
9. Wang H. Proxy provable data possession in public clouds. IEEE Trans Serv Comput. 2012; 6(4):551–9.