A Study of Data Security Integrity in Cloud Computing

Thejoramnareshreddy Boya[1], Dr. Neeraj Sharma[2]

[1]Research Scholar of Sri Satya Sai University of Technology & Medical Sciences, Sehore, M.P., India.
[2]Research Supervisor of Sri Satya Sai University of Technology & Medical Sciences, Sehore, M.P., India.

Abstract

Cloud can be envisioned as a working model for providing an environment which enables the cloud users to access a shared pool of virtualized and configurable computing resources that can be easily provisioned as services on the pay-per-use manner from the cloud service providers. Amongst the services, the cloud storage services gain more attraction due to the growth of digital data at an incredible rate. By utilizing the cloud storage services, the cloud users can be relieved from the burden of local data storage and maintenance. But, the users are unaware of the exact location of the remotely stored data. This unique feature imposes many security challenges like data theft, damage of information, disruption of services, loss of privacy, etc., on the data stored in the cloud. In order to alleviate the challenges, this research aims at developing model for cloud computing so that the data stored in the cloud are confidential, integral and authenticated. Cloud users depend on cloud for storage of massive amount of data. Before storing the data, the data owner should imply the proper security mechanism that ensures confidentiality, integrity and authentication in order to protect the stored data from unauthorized access. In cloud, confidentiality of data can be compromised by applying the attacks like data leakage, phishing and eavesdropping. To overcome this, this research aims at developing a new encryption algorithm called 2-keys symmetric encryption algorithm that can be used for the encryption of massive data. This encryption procedure incurs lesser computation overhead and minimum time for performing encryption and decryption.

Keywords: *Data Security Integrity, Cloud Computing, Data Storage, Encryption Procedure.*

1. INTRODUCTION

Data integrity is ensured by generating the metadata for the data to be stored in cloud by utilizing the position of individual characters of the data blocks. This metadata is attached to the encrypted data blocks and is stored in the cloud storage servers. Then, the originality of data can be verified by the data owner or by some Third Party Auditor (TPA) by comparing the encrypted and metadata blocks. Security of computation done by the Cloud Service Provider (CSP) is ensured by using the concept of Merkle hash tree. The approach for data and computation integrity is used to eliminate the attacks like salami attack, Byzantine failures and omission of rarely accessed data by the CSPs. Moreover, the experimental analysis proves that the data integrity and secure computation are ensured with reduced computational and storage overhead of the cloud users. The secured storage for dynamically changing data stored in the cloud is achieved by using the well-known data structure called linked list. This linked list is generated and maintained by the data owners. Using this scheme, dynamic data operations like insertion, modification, appending and deletion are possible without unnecessarily downloading the whole data from the cloud server. Based on the experimental results, it is proved that this method detects the corruptions or modifications on the dynamically changing data with the highest probability.

This research also constitutes a remote user authentication mechanism for the cloud users by using the fingerprint of data users as an authentication structure. An interesting feature here is that, it is not necessary for the cloud users to submit their fingerprint as such to the cloud service provider, thereby preventing the misuse of sensitive user authentication details. Only the signature generated from the hash values of the fingerprint using the concept of Merkle hash tree is submitted to the cloud provider. This remote authentication technique is used to eliminate the attacks like password guessing attack, dictionary attack, spoofing attack and Denial-of Service (DOS) attack. Here, it is also proved that the performance of the remote fingerprint authentication scheme in terms of accuracy, error rate, storage, computation and time complexity is high.

Message Authentication Code (MAC): To verify the data integrity, MAC for the entire data is generated by the DO before storing the data file in a remote server. It is retained by the DO in the local storage, but the original data is stored in the remote server. In order to verify the integrity of the data, the data owner retrieves the entire data from the remote server, re-computes the MAC and compares it with the one that is retained in the local storage. By comparing the equivalence of the two MACs, the DO confirms the accuracy of the data stored in the server. The drawbacks of using this MAC are the unnecessary computation overhead in calculating MAC many times and unwanted bandwidth in transferring the whole data file from the remote server to the cloud user.

Checksum: Before storing the data file in the cloud server, the data user computes the checksum of the data file and retains the checksum in the local storage. To verify the data file, the data owner retrieves the data from the server and computes the checksum. If this checksum matches with the checksum in the local storage, the data user ensures the integrity of the data file.

Provable Data Possession (PDP): The model of Provable Data Possession scheme [86] assures the integrity of data files stored in the cloud servers. This scheme uses RSA based homomorphic tags for verifying the data integrity. These tags should be pre-computed by the data owner and later used for verification. The drawback of this scheme is a computation and storage overhead in generating and maintaining the tokens.

Pre-computed tokens: Before storing the data in the server, a number of verification tokens each covering a portion of data blocks are computed and retained by the data owner in the local storage. To verify the correctness of the data stored in the cloud server, the cloud user challenges the server indicating the position of data blocks. For that, the cloud server generates the signature on the indicated data blocks and sends it to the cloud user. By comparing the signature with the token in the local storage, the cloud user decides on the integrity of the data blocks. The drawback of this scheme is that, the DO has to compute and store more number of verification tokens in order to ensure the integrity of the entire data incurring heavy computation overhead.

Proof of Retrievability (POR): This scheme uses some sentinel characters to check the integrity of the data file stored in the cloud server. These sentinel characters are hidden among the original data blocks by the data users. The positions of the sentinel characters are maintained in the local storage by the data owner. For checking the integrity, the data owner sends the challenge message specifying the position. The server responds with the corresponding character in the requested position. Then the server decides on the integrity of the data block by matching the characters.

2. SECURITY OF SERVICE DELIVERY MODELS IN CLOUD COMPUTING

Due to the specifics of each service delivery model in cloud technologies, the security is considered according to SLAs between the user and the provider for each of the following three main models:

1. Software as a Service

With this model, users pay a subscription for a software product, whereby part of the information or the full information is saved on a remote location and users can access this service via the Internet. With SaaS, users do not have any control or authority to modify or manage cloud infrastructures or even specific applications that are already developed. Users have limited options to configure settings related to the use of these applications. The provider fully controls the cloud infrastructure and is responsible for the confidentiality, integrity and accessibility of the data and information.

2. Platform as a Service

This model enables users to develop their own cloud infrastructure applications using programming languages and additional software tools, which are provided by the cloud service provider (such as .NET, Ruby, Java, etc.). PaaS provides users with all the resources they need, so as to be able to develop applications and services entirely on the Internet without the need to download or install additional software. The user is not yet able to manage the underlying cloud infrastructure but only the applications developed by him/her. One disadvantage of PaaS is the lack of interoperability and transfer of applications, which are already developed by the user, to other providers. I.e. if the user wants to transfer the applications developed with the current provider to another cloud provider, this cannot be done or if it can be done, the costs will be extremely high. Another disadvantage is that if the provider decides to

leave the business, the applications and information in them will be deleted. With this model, the providers' liability is associated only with the integrity and accessibility of data. The user is responsible for the confidentiality and protection of the information.

3. IaaS

The infrastructure as a service allows the respective organization to create its own software environment. In the SaaS and PaaS models, the provider provides the user with applications but in the IaaS model, this is not done. This model only provides the hardware on which the organization can install whatever it wants. The control, exercised by providers, is at an extremely low level. They are only responsible for the accessibility of the services provided by them. Compared to providers, the users' responsibility is at a quite high level. They are responsible for the confidentiality, integrity of the data and its protection. The following table shows the responsibilities of cloud service providers and users for each model. Before an updated version of a data security model in cloud technologies is proposed, the evolution of cloud technologies should be known very well, as well as their advantages and disadvantages. The following figure shows the overall development of these technologies by years.

3. NEW DATA SECURITY MODEL IN CLOUD COMPUTING

The development of the new data security model in cloud technologies and related service delivery models is based on the CIA triad for cloud services, which is described above. The CIA shows the responsibilities of providers and users for each of the three most widely used cloud service delivery models. For example, in the case of the most widely used model globally – the Software as a Service, privacy, data integrity and availability are the sole responsibility of the provider of cloud services or products. For the other two models – the Platform as a Service and Infrastructure as a Service, the responsibilities are allocated differently between the user and the provider. In this allocation, the most significant problem with cloud technology remains - the problem of data security and protection. The new data security model aims at increasing the level of security of data and programs by checking users with the so-called three-factor authentication from another server. Three-factor authentication uses a combination of the following methods:

- **Something the user knows:** it is a symbolic sequence that is known only to a given user and is unknown to others. When it is claimed with the user ID, it is taken as proof of possession of the identity. Typical examples of something known are passwords, PINs, private keys, etc. The most common and accessible means of authentication are passwords, which are a sequence of characters chosen by the system administrators or by the users themselves.

- **Something the user owns:** it is a material object (card or other medium) on which additional information is usually recorded. In the process of authentication, the user claims the owned object by placing it in a device that can read its contents and confirm or reject the declared identity. Such techniques, for example, can help authenticate users using electronic signatures to share information with different institutions.

- **Biometric data:** it represents unique data that can belong to only one individual. Fingerprints, iris, retina, voice recognition, etc. can be included in this category.

The model achieves a higher degree of security, as it focuses mainly on the preliminary control – the highest degree of efficiency. SSL or IPSec, as well as DoS protection, are added here. The model also uses the so-called Single sign-on, which also requires authentication from an external server. Once the user is admitted to the system (cloud), he/she can access all cloud infrastructures in the environment (private, public, and community) through Single sign-on without the need to enter authentication data again. That is, once a user is successfully admitted to the cloud infrastructure, this means that he or she has successfully passed the three-factor authentication described above and has the right to use his or her access to enter other cloud infrastructures in the environment, without the need to authenticate to each cloud individually (as it will be the case without a Single sign-on).

The model encrypts the information before it sends it to users by SSL, IPSec protocols. Another improvement of the model is that authenticated users and end users cannot be eavesdropped, which guarantees a high degree of confidentiality of the model. The model is mainly used in a cloud environment that requires maximum security and protection of data and information and enables the successful collection of biometric data. Cloud providers are primary responsible to secure the information. Each provider uses special techniques to secure its resources in this

system. Upon entry of a username and password, Single sign-on also allows some authenticated users to use this registration for other sites and applications.

4. SYSTEM MODEL FOR DATA INTEGRITY ASSURANCE

This system model for secured storage consists of various entities like Data Owners (DO), Data Users (DU), Cloud Service Providers (CSP) comprising of n cloud storage servers (s1, s2,.. sn) and a number of compute intensive servers, a separate service for generating meta-data and a Third Party Auditor (TPA).

The role of the auditor falls into two categories.

1. Private auditability: In this, only the data owner is permitted to check the integrity of the data file residing on the server.

2. Public auditability: It means that anyone, including the TPA is permitted to verify the data integrity.

Metadata generator: It is an entity or a service that is actually running on the premise of cloud service providers, but under the control of the Data Owners. It is performing the task of generating the metadata from the encrypted data blocks.

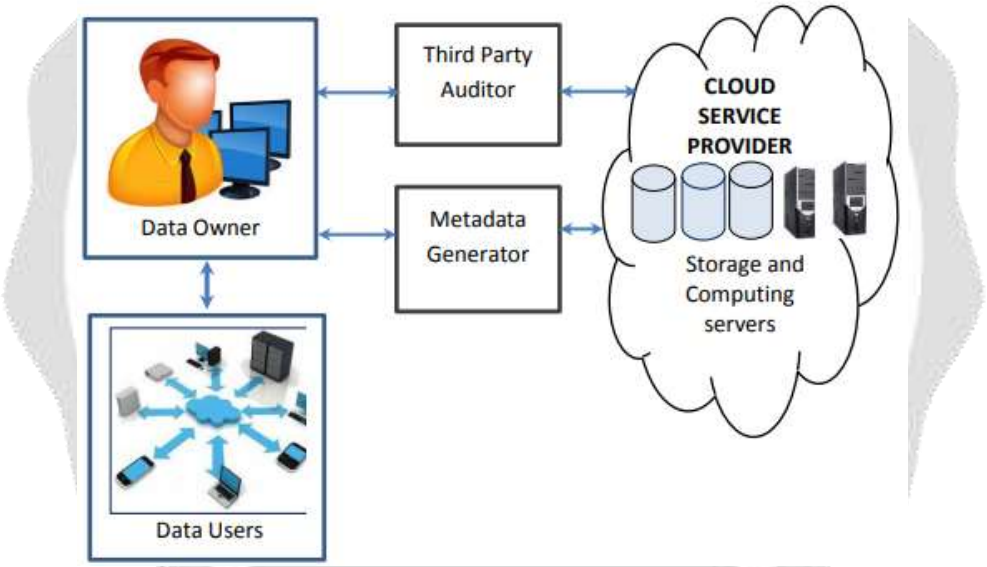

Figure 1: System model for data integrity assurance

5. FUNCTIONS FOR ASSURING DATA AND COMPUTATION INTEGRITY

The aspect of integrity assurance for the cloud storage is twofold. The first one deals with a mechanism for checking the integrity of the data, while the second provides an approach for verifying the correctness of computations done by the cloud service provider.

1. Data integrity assurance

In this, the integrity of the outsourced data stored in the un-trusted remote cloud servers has been assured. It has been done by implementing a method that acquires a proof of data possession by generating metadata of the data in the cloud. This proof verifies that the data stored in the remote cloud server are not modified by unauthorized users, thereby assuring the data integrity. So, this verification protocol prevents the remote cloud storage servers and unauthorized individuals from damaging, misrepresenting or changing the data without the knowledge of the data owner by conducting frequent security checks on the data storage.

2. Computation security assurance

The cloud users depend on the CSP for performing some computational tasks. To accomplish this data users submit the CSP, the data on which the operations have to be done. The CSP performs the computation and submit the result to the cloud users. In doing these computational tasks, the cloud server may elude from doing the computations mentioned by the cloud users in the following ways:

- The cloud server applies some easy random functions and returns the cloud users a random result or the result of previous computations instead, but claims to have done the correct computations.
- Instead of applying the original data, the cloud service provider may apply the incorrect data which incurs a lesser computational cost and claims to have applied the correct data.

6. ASSURING THE INTEGRITY OF DATA AND COMPUTATION

For ensuring the correctness of users' data stored in the cloud server, an effective scheme has been proposed here. This scheme benefits with two salient features.

- Assuring integrity of the data stored in cloud server by obtaining a proof that the data is not modified or deleted by the cloud service provider without the consent of the data owner.
- Ensuring secure computation by allowing the data users to verify the correctness of computations done by the cloud service provider, thereby avoiding the misbehavior of the server.

Metadata generation phase

1. The data file is split into nb data blocks by the data owner.
2. All the data blocks have been encrypted using the 2-Keys symmetric encryption algorithm by the encryption service, thereby producing the encrypted data blocks EDBs.
3. Then the EDBs are given as input to the metadata generator.
4. The metadata generator applies some random function RF and the secret key Sk, received from the data owner over the encrypted data blocks EDBs and generate metadata blocks MDBs.
5. These MDBs are attached at the end of the encrypted data blocks EDBs by the metadata generator.
6. The combination of EDBs and MDBs are stored together as a single data file in the cloud server by the metadata generator.

Computation commitment generation phase

1. Cloud users submit a number of computational tasks along with the data to the cloud service provider.
2. On receiving these, the CSP performs the computations and generates the signature using the Merkle hash tree by considering the results of computations as the leaves.
3. Then the CSP sends the signature and the computation results to the cloud user.

7. CONCLUSION

After the massive implementation of cloud technologies in businesses and enterprises, the security of users' information and authentication has become increasingly important issue, which is discussed by many organizations around the world. Because of security issues, many users and businesses still refuse to use the cloud-based model. Risks, threats and vulnerabilities in virtual environments of cloud technologies significantly differ from the ones in physical environments. This publication introduces a new CIA triad data security model. It adds a three-factor authentication (3FA), as well as a Single sign-on using the Open ID standard. Thus, the effectiveness of preventive control is increased – this is the most important and effective control in an organization. The model focuses on data security and protection. Security management in an organization should also be in line with IT policies and standards for cloud technologies, as well as with business objectives for data security through the CIA triad – confidentiality, integrity and availability of information. The proposed model will significantly increase data security in cloud technologies and thus reduce the risk of misuse and misuse of personal data, as well as the misuse

of identity. Cyber threats and cyber-crimes will be reduced. The use of smart technologies will be safer and psychological stress and distrust in consumers will be reduced.

8. REFERENCES

1. El-Gazzar R., Hustad E., (2016), Olsen D, Understanding cloud computing adoption issues: A Delphi study approach, Journal of Systems and Software, Volume 118, pp. 64-84. https://doi.org/10.1016/j.jss.2016.04.061
2. Viega J., (2009), Cloud Computing and the Common Man, in Computer, vol. 42, no. 8, pp. 106-108, DOI: 10.1109/MC.2009.252
3. National Institute of Standards and Technology(NIST), U.S. Department of Commerce, The NIST definition of Cloud Computing, Special publication of Computer Security, 2011
4. Center Of The Protection Of National Infrastructure CPNI by Deloitte, "Information Security Briefing of Cloud Computing"; The Security of Cloud Services, A Middle East Point of View, 2017
5. Ajoudanian Sh., Ahmadi M., (2012), A Novel Data Security Model for Cloud Computing, IJET 2012, Vol. 4 (3): pp. 326-329, DOI:10.7763/IJET.2012.V4.375
6. Mohamed M., Abdelkader H., El-Etriby S., (2012), Enhanced data security model for cloud computing, 2012 8th International Conference on Informatics and Systems (INFOS), Cairo, 2012, pp. CC-12-CC-17.
7. Hoyer U., Obel H., (2017), Guide on SaaS vs PaaS and IaaS, https://www.linkedin.com/pulse/guide-saas-vs-paas-iaas-ulrik-hoyerhansen-obel
8. Garvanova M., Shishkov B., Janssen M., (2010), Composite public values and software specifications. Business Modeling and Software Design. BMSD 2010. Lecture Notes in Business Information Processing, vol. 319, 412-420. Springer, doi: https://doi.org/10.1007/978-3-319-94214-8_32.