IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 16, Issue 5, Ver. IX (Sep-Oct. 2014), PP 71-76
www.iosrjournals.org

Analysis on Data Integrity in Cloud Environment

Siddhartha Rao, Savan Gujrathi, Mithun Sanghvi, Shubham Shah
(Dept. of Computer Engineering, Vishwakarma Institute of Information Technology, Pune, India)

Abstract: *Cloud Computing privileges convenient, easy on-demand access to a collection of configurable resources and services. The services and resources on the cloud can be robustly deployed with low maintenance cost, effort and high efficiency. Cloud is still considered as an insecure computing platform from the user point of view due to the lack of security in terms of confidentiality and integrity. Shielding mechanisms for cloud systems must be designed that safeguard sensitive data using cryptographic techniques and also maintain the intactness of user information stored on the remote storage by protecting it against malicious entities or behaviors. In this paper, we enunciate an overall sketch of various integrity issues in cloud that have been discovered over a period of time and the possible solutions that deal with these serious issues.*

Keywords: *Cloud Security, Data Integrity, Security, Integrity Verification, SHA-1*

I. Introduction

It's been a while since we have been surrounded by the concept of Cloud and the technology. The term "Cloud" itself is so funny that it flashes a literal picture of the cloud in the first place. Put in simple words, it is a giant clusters of servers consolidated into one big data bank located at different places working to offer various services on the fly.[4] Cloud Computing is looked upon as a technology that will help in reducing maintenance and development expense, in contrast to yield high performance services. Cloud Computing is an emerging trend and is also mutating at rapid rate. This technology assures high cost benefits. Information and applications are stored on cloud servers which are known as Data Centers. Data Center Environment helps organizations to execute programs faster, with simpler conformity and lower management overheads, and swiftly scale resources to incorporate constantly changing industrial needs. The Data Center stores client information that usually would have been stored on the client machines. This evokes worry in relation to client's data privacy as the client has to outsource the information. The outsourcing of data to centralized cloud services can influence the security and privacy of the end-user. Using of centralized virtual framework might harbinger advanced attacks to the data of the user. The overall global usage of cloud makes possible the resource optimization possible on a wide scale.

Cloud computing is a flexible, efficient & scalable method to organize resources like software, infrastructure, storage.[3] It provides user services in various different ways, suiting the user's need in terms of bandwidth, network-access, applications etc. The primary objective of this paper is to provide security in terms of integrity of the client's data, for a substantially large amount of time (in years), which is stored on the cloud. We describe the definition of data integrity in Section II, followed by the objectives in Section III and some of the issues and challenges in Section IV. Further, we discuss some probable solutions in Section V and lastly conclude in Section VI.

II. Definition Of Data Integrity

Data integrity in simple terms can be understood as the maintenance of intactness of any data during transactions like transfer, retrieval or storage. To make a child understand the meaning of data integrity, it can be defined as ensuring that the data is unaltered, correct and consistent. The data may change if and only if an authorized operation is valid on the data.[3] Integrity of the data can be hampered at any level of storage, any type of factor being the reason. Therefore, for the same reason, integrity surveillance in cloud storage is the most critical issue for any data center. The examples of different media types that can cause the loss of integrity are bit rot, controller failures, metadata corruption, duplication and tape failures. Bit rot is the most critical among them as integrity is affected even if any bits of data are also altered by any reason. For instance, integrity of a text file can be affected by adding just a space character in the file. In such cases, altering of few words could be highly risky. One more reason for data corruption can be the migration to different platforms. Cloud storages are data centers, and are still vulnerable to data corruption even though it has been in use since years now. Clouds are trusted for data storage; security in terms of integrity is the paramount aspect that is focused in this paper.

One of the most popularly used method to verify the integrity of any data is based on the hash values. By implementing a pre-defined algorithm, a hash value (known as the fingerprint), which is unique is derived for a set of data and it is always the same value for that data until it is altered by any means. Since the hash value corresponding to a particular data never changes, if a copy of that data, which is stored on the Cloud, is retrieved after say 2 years and if the two hash values are not identical, it is an indicator that at least one of the two copies has been either altered or corrupted.

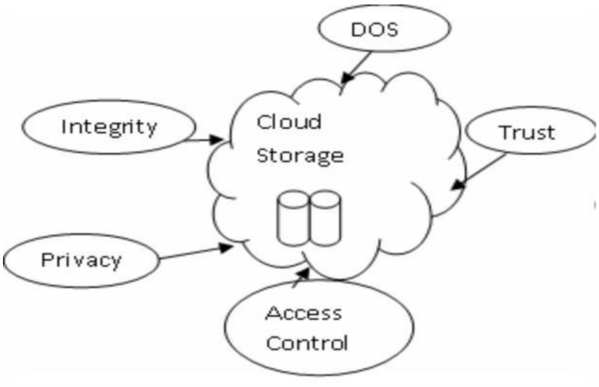
III. Objectives

Cloud service provider could unintentionally or purposely modify or erase any data from the cloud server.

Therefore, there should be some kind of system to assure the integrity of the data. The ongoing Cloud security model is positioned on the inference that the client should believe the provider. It is generally administered by an agreement known as Service Level Agreement (SLA). SLA defines notions that are bilateral to the client and the service provider. The objective of this study is to present an analysis of the existing cloud security systems requirements and highlight the existing vulnerabilities and possible threats in effective security schemes for cloud systems. It will surely throw light on security issues and solutions that will help researchers to identify requirements at different levels and recognize threats in various cloud computing models eventually leading to enhanced improvement in the security domain of cloud.

IV. Privacy Issues And Challenges

The rapid growing "Cloud computing" technology is globally used in many sectors due to its multipurpose services offered at low cost. For instance, pay-per-use storage, where the user pays only for the amount of storage space demanded on the cloud via the Internet. Also, there are on-demand type of services like the retrieval of any data and synchronizing it over various devices.



Security issues in cloud

Clouds also provide remote processing of data like file conversions instead of accomplishing the task on the client machine.[5] Cloud storage is one of the services offered to store user's data on the remote server. Even though the cloud providers guarantee that the users' data will be secure and consistent, there are many problems and challenges that still prevail in the security domain of the cloud that need to be dealt with. Data safety in the form of integrity is the chief concern in cloud environment.

Cloud computing provides 3 major categories of services:

1. Infrastructure as a Service (IaaS)
2. Platform as a Service (PaaS)
3. Software as a Service (SaaS)

Systems incorporating cloud storage services must make sure that their data is highly secure before, during and after transmission of it, as the user's control is forfeited once it reaches the cloud. However, there is no assurance that a cloud vendor will secure the data and preserve its integrity. A user also has no knowledge about the measures a vendor has implemented so as to trust him. There is always the feeling of insecurity in this case. Users must be capable to examine the remote cloud system and be assured about their data privacy throughout all phases right from uploading to storing. In general, it is difficult to deal with the issue of an untrusted host when interacting with a remote system as it can eavesdrop on the data in the uploading phase. The flexibility and virtualization principles of cloud contribute to such threats. Neglecting the need of integrity at data level could result in heinous problems.[2]

Some of the identified major cloud security issues include:

1. **Virtualization**: The cloud computing runs on the technique of virtualization where same system is used by multiple users which may lead to data un-safety.

2. **IP address**: If the same IP address is used by different users then it may lead to access of different data of other users.

3. **Elasticity**: Many different users operate the same machine to run their application which may lead to data breach.

4. **Network Insecurity**: Due to the large amount of traffic and uncontrolled nature of the network there may be loss of data.

5. **Insecure APIs**: Implemented APIs are a common source of malicious infections. Insecure or deprecated APIs may adversely access data

6. **Provider Security Malfunction**: Lack of cloud provider security due to weak security planning and implementation could cause the database to compromise.

7. Reliability and Availability Issues

8. Native Customer Attacks.

IV.I Integrity Vulnerability

Consider the example of a Google App Engine (GAE); there are a series of phases through which any user request transitions. Let us see the vulnerability in this simple scenario. When a user wants to retrieve some data, he sends an authentic data request to the cloud. Similarly when he wishes to store some data, he again sends a request for the same. But at the same time an MD5sum or an SHAsum is also sent along with it. When the service provider receives it, he stores that data and the associated sum (hash value). When the provider receives a valid authentic data retrieval request from any other or the original user, it will transmit the data with the sum. The integrity during sending can be maintained by SSL channel. However, from the cloud storage service point of view, the maintenance of data integrity does not just depend on securing the uploading and downloading processes, but also depends on safeguarding data in the storage medium. There are various techniques developed like using a Trusted Third Party Auditor or Encrypting whole of the database along with its field names etc. The upload process can only ensure data sent to the cloud is what user sent and the download process can only ensure data received by the user is what the cloud server sent. Unfortunately, this procedure cannot guarantee the data is not modified in the storage space.

Thus in this scenario three important concerns raise:
(1) **Confidentiality**: Even though a cloud vendor (X) has full access to all user data (Y and Z), X is considered as malicious/untrusted so Y and Z do not want to expose their data.

(2) **Integrity**: Being the admin, X has full power to play with all the data. If Y is trying to retrieve some data that is sent by Z on the cloud, what guarantee does Y have that the data is the same as sent by Z or tampered by X?

(3) **Repudiation**: In case Y finds out that the data has been played with, is there any evidence for proving X guilty? Similarly, X also needs some evidence to defend himself and prove his innocence.

V. Solutions

Ensuring data integrity requires a bond of trust between the client and the provider. The conspicuous factor in ensuring the integrity of the generated data is having trust on the process generating that data. Also, the input given to the process plays an important role. [7] Some of the solution techniques to assure integrity of client data are:

- Semantic examination (integrating logic in the process of semantic analysis)
- Certificate (authorized proof from trusted agencies)
- Trusted route (ensuring the data travels via an authorized channel)

Many cryptographic techniques have been devised and deployed to provide better security for the sake of protecting integrity factors. Provision of secure communication channels and execution services to prevent the interception of sensitive information is highly recommended. Evaluation [6] of integrity requires checking of the source of data, individuals provided access to data, calibration of measurements and knowledge of fingerprints (MD5, SHA-1, SHA-256, SHA-512, etc.). The most popular existing technology is the checksum that is already in force especially at times where user tends to download files. For instance, on the eclipse website, they have made available an MD5 or SHA checksum alongside every file that the public desires to download. When the user downloads any file, he can generate a checksum on his machine and compare it with the hash that eclipse has provided. If both the hash keys match, it indicates the intactness (integrity) of this file, else the file has to be re-downloaded. Few more solutions that should be made effective by the cloud providers for any of the discovered issues are:

1. Using Data Encryption techniques [5]

First solution is providing a good data encryption standard. For more security purposes, data encryption burden for enhanced security should be bore by the cloud providers so that no separate security from the enterprise is required. IT leaders and developers should device a strategic implementation of finding the probable flaws and to determine where the encryption is required.

2. Better Enterprise Infrastructure

Enterprise should possess and infrastructure that enables and facilitates the installation of software compo nents such as thin clients, OS and hardware essentials like firewalls, proxy servers etc. These enforcements should be capable to restrain cyber-attacks.

3. Good Recovery Schemes

Cloud providers should provide good recovery schemes in case of any scenario of data lose or fragmentation. In this way, the data can be recovered and consistent continuity of data is possible.

4. Right Provider [5]

One of the most important ways is to find the right cloud vendor. Different vendors have their customized different data management schemes. A cloud vendor should be renowned and established, should have a good experience. In this way, the chances of a cloud vendor closing are minimized.

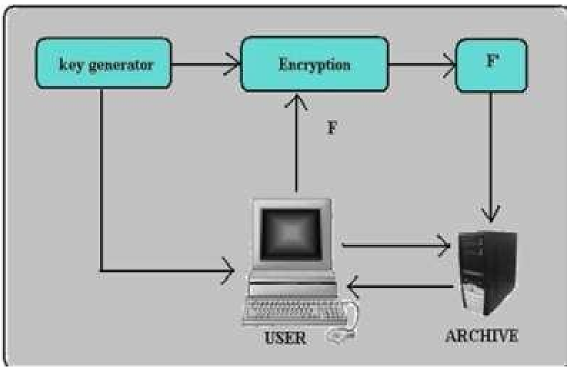
5. Good and Clear Contract Documentation

A contract between the vendor and the enterprise should be made clean and clear. In any scenario if the vendor shuts his services against the terms specified in the contract, the enterprise can claim.

V.I Proposed techniques for ensuring Integrity of Data

1. Proof of Retrievability (POR) [1]

Juels and Kaliski proposed a scheme called Proof of Retrievability (POR). POR scheme says that verifying the data stored by the user on a remote cloud system is not altered by the cloud itself. POR for colossal file sizes are called as sentinels. The primary act of sentinels here is allowing the cloud to access only a part of a file instead of accessing the whole file. Sravan and Saxena proposed a Schematic view of a Proof of Retrievability based on inserting random sentinels in the data file.



Schematic view of a POR

The architecture shows that the client (user) stores his file (F) on the cloud server (archive). But before he stores any data on the cloud, he must encrypt the data in order to avoid any malicious attack or unauthorized accesses hamper it. Now the user can request the server anytime, maybe in one year or maybe in ten years according to his wish. As soon as the user sends a challenge to server in order to examine the integrity of his file (F), the server will ask to return a set of sentinels in the challenge response. If the data is deleted or corrupted or modified, the sentinels may get corrupted and hence the server fails to prove the integrity of the user's data. In this way, the client can gain assurance via proof that the server has corrupted, modified or lost the file.

2. Provable Data Possession (PDP) [1]

A Provable Data Possession scheme makes sure that the files in the cloud server are retained. The owner of that data processes his files to store its metadata locally. The file is then uploaded by the owner to cloud and deletes its local copy. The owner checks that his data is contained in the cloud by using the challenge response protocol. In this way, the clients use this integrity proving technique to verify the integrity of their data and to periodically check its availability. Therefore, this technique guarantees cloud security to the client.
Limitation: Using this method, a sure proof is attained that the file (F) is possessed by the server. It has a high computational overhead due to which the computation cost is extremely high. The overhead increases as the hashing process is run on the entire file every time the client attempts to check its integrity.

3. Using Strong Algorithms [2]

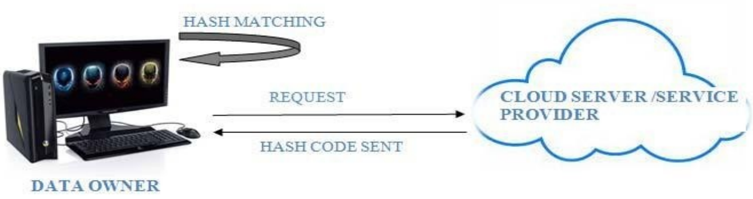
We assume that our file F is divided into n blocks. Each block is encrypted using Advanced Encryption Scheme (AES) algorithm. There are two fields for the data to be encrypted, public key and private key. The data is encrypted using public key and decrypted by the private key which is known only to the owner. After the encryption a hash key is generated using a Secure Hash Algorithm (SHA). Then, for each encrypted hash codes the digital signature is attached to ensure authentication. Following are the steps to be carried out by data owner while file being stored in the cloud.

Step 1: File F is divided into n blocks.
Step 2: Encrypt each block using encryption scheme.
Step 3: Generate hash codes for each encrypted block using secure hashing. Step 4: Digital Signature is attached
Step 5: Copy of ID, Hash Code, meta data, time stamp are stored in the data owner side.

4. Integrity Verification for Static Data [2]

The data owner can verify the integrity of the outsourced data by requesting the server to send the hash code for any block. The following steps help to check the integrity of the data, by matching the hash codes of the data that is stored with the one which is requested from cloud server.

Step 1: Generate the random set
Step 2: Request the hash code for the generated random set.
Step 3: Match the retrieved hash code with the one already in the data owner side. Step 4: If hash code, timestamp and size match then integrity is verified.



Integrity Verification

VI. Conclusion

Every coin has two sides. Even though the Cloud technology offers great potential to improve quality, optimize productivity and is cost effective, it also includes many security challenges related to the data safety and accuracy. As the cloud is mainly used for storage purposes, data integrity is of primary concern to the user as the control over the data is relinquished from the user. There are many methods available to us out of which Provable Data Possession (PDP) and Proof of Retrievability (POR) have been well analyzed. The past literature works prove that some of the methods facilitates the user in having a proof of integrity of the data that the user stores on the cloud with trust. The study done in this paper will help the researchers to take these issues and solutions into detailed discussion giving then a new view point. It still cannot handle the case when data needs to be changed dynamically thus researching and developing on this will be a future challenge.

Data integrity being a necessary safety in all types of storages, making it applicable for malicious and unreliable third party hosts is another future challenge.

References

[1] Vitthal Raut, Prof. Suhasini Itkar, A Survey on Data Integrity of Cloud Storage in Cloud Computing, IJAFRC Feb. 2014, Vol1 Issue2, pp. 58-64.

[2] V.Nirmala, R.K.Sivanandhan, Dr.R.Shanmuga lakshmi, Data Confidentiallity and Integrity Verification using User Authenticator scheme in cloud, in Proc of ICGHPC Mar. 2013.

[3] Rajkumar Chalse, Ashwin Selokar, Arun Katara, A New Technique of Data Integrity for Analysis of the Clou d Computing Security, CICN 2013, pp.469-472.

[4] Venkatesa Kumar V, Poornima G, Ensuring Data Integrity in Cloud Computing, IJCA Feb 2012,Vol5 Issue4, pp.514-519.

[5] Pradeep Kumar Tiwari, Dr. Bharat Mishra, Cloud Computing Security Issues, Challenges and Solutions, IJETAE Aug. 2012, Vol2 Issue8, pp. 306-309

[6] Jun Feng, Yu Chen, Wei-Shinn Ku, Pu Liu, Analysis of Integrity Vulnerabilities and a Non-repudiation Protocol for Cloud Data Storage Platforms, Manu. Submitted to SCC, 12 Jun. 2010, in conj. with ICPP 2010, pp. 1-8.

[7] Jeff Naruchitparames and Mehmet Hadi Gunes, Enhancing Data Privacy and Integrity in the Cloud, unpublished.