Identifying Data Integrity in the Cloud Storage

Saranya Eswaran[1] and Dr.Sunitha Abburu[2]

[1] Adhiyamaan College of Engineering, Department of Computer Application, Hosur.

[2] Professor and Director, Adhiyamaan College of Engineering, Department of Computer Application, Hosur.

Abstract

In cloud computing, data is moved to a remotely located cloud server. Cloud faithfully stores the data and return back to the owner whenever needed. But there is no guarantee that data stored in the cloud is secured and not altered by the cloud or Third Party Auditor (TPA). In order to overcome the threat of integrity of data, the user must be able to use the assist of a TPA. The TPA has experience in checking integrity of the data, that clouds users does not have, and that is difficult for the owner to check. The data in the cloud should be correct, consistent, accessible and high quality. The aim of this research is twofold 1) ensuring the integrity of the data and provides the proof that data is in secured manner. 2) Providing Cryptographic key to secure the data in the cloud. The proposed approach is been implemented and the test results are promising.

Keywords: *Data integrity, Cryptography, TPA, Cloud storage.*

1. Introduction

Cloud storage is visualized pools where data and applications are stored which are hosted by the third party. Company, who desires to store their data in the cloud, buy or lease storage capacity from them and use it for their storage needs. Some of the cloud storage benefits are reduce costs, provide more flexibility, reduce IT management of hardware and data, reduce management of web applications through automated updates, and provide greater storage capacity. In spite of these benefits, "cloud" lack in some of the issues like data integrity, data loss, unauthorized access, privacy etc.

Data Integrity is very important among the other cloud storage issues. Because data integrity ensured that data is of high quality, correct, consistent and accessible. After moving the data to the cloud, owner hopes that their data and applications are in secured manner. But that hope may fail some times (i.e.) the owner's data may be altered or deleted. In that scenario, it is important to verify if one's data has been tampered with or deleted. To validate data, often a user must download the data. If the outsourced data is very large files or entire file systems, such downloading to determine data integrity may become prohibitive in terms of increased cost of bandwidth and time, especially if frequent data checks are necessary. This paper propose a method that, owner need not download the data or files to check the integrity and it provides the proofs that data is stored at a remote storage in the cloud is not modified by anyone and there by integrity of the data is assured. Some of the best examples for cloud storage are Amazon S3, Windows Azure Storage, EMC Atmos, FilesAnywhere, Google Cloud Storage, Google App Engine Blobstore, iCloud by Apple.

The remainder of the paper is organized as follows: Section two analyses about the cloud storage architecture and along with its characteristics. Section three of this paper briefly describes the proof of Retrievability and role of Third party auditor (TPA). Section four is explaining how the data integrity is verified in the cloud. We concluded the paper in section five.

2. Cloud Storage

The process of storing data in the remotely located cloud servers are said to be cloud storage. The Architecture of cloud storage as shown in fig. 1.

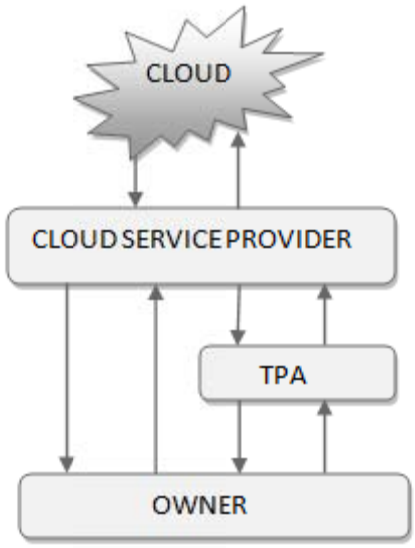

Fig. 1 Architecture of cloud storage

The cloud storage is better than all traditional storage methods. Because of the following reasons

- ✓ Companies do not need to install physical storage devices in their own datacenter or offices.
- ✓ Storage maintenance tasks, such as backup, and purchasing additional storage devices are offloaded to the responsibility of a service provider, allowing organizations to focus on their core business.
- ✓ Companies need only pay for the storage they actually use.

Some of the characteristics of cloud storage are as follows:

Performance: In the internet, TCP can controls the flow of data based on packet acknowledgements from the prior endpoint. But TCP is supreme for moving small amounts of data through the global Internet and it is not suitable if size of the data becomes larger. Cloud storage overcomes the above problem effectively.

Manageability: Basically the main key focus of the cloud storage is maintenance. Clients itself store their data with them but it is too expensive when the data size is increased and also maintaining their data is not an easy task. For this reason, cloud storage must be self-managing to a large extent. Client stores their data on the cloud by this maintenance trouble is reduced.

Availability: In Cloud storage, data are retrieved frequently, quickly and securely.

One way to protect against threats to web applications and data is to deploy a Web application Firewall as a software solution. No additional hardware is required on the part of the cloud provider and it can be installed directly in web applications. When deployed correctly, a Web Application Firewall protects web applications and data from known threats including Path Traversal, Remote Command Execution, and compromised servers. But the firewall must consume CPU cycles reading for each packet, this process requires more processing power, which become a bottleneck for the network. This means application firewalls are less suited to real-time applications. But cloud storage is well suited for accessing huge size file and complicated real-time applications.

3. Literature Review

In cloud computing, enormous threats are raised. One of the threats is data privacy and integrity. A lot of researchers focused on proving data integrity in the cloud and introduce many solutions to decrease the threat of the data privacy and integrity. Calce[1] says about cloud computing, putting everything into a single box will only make it easier for hackers. Moving to a virtual environment to save on costs automatically introduces fresh risk on top of existing risk. Priya Metri and Geeta Sarote [2] introduce threat model to treat the privacy problem in the clouds. One of the service is third party auditing because it notify the threats in cloud computing is tampering with data in the cloud that interfere with the unauthorized modifications for the data, which lead to an effectiveness processors, data storage and data flow. Proofs of Retrievability(POR) model proposed by Juels and Kaliski [3] are among the first few attempts to formulize the notion of “remotely and reliably verifying the data integrity without retrieving the data file”.

Archival network storage [4] presents unique performance demands. File data are large and are stored at remote sites, accessing an entire file is expensive in I/O costs to the storage server and in transmitting the file across a network. Reading an entire archive, even periodically, greatly limits the scalability of network stores. Furthermore, I/O incurred to establish data possession interferes with on demand bandwidth to store and retrieve data . Previous solutions do not meet these requirements for proving data integrity. Some schemes provide a

weaker guarantee by enforcing storage complexity moreover, all previous techniques require the server to access the entire file, which is not feasible when dealing with large amounts of data. This paper conclude that clients need to be able to verify that a server has retained file data without retrieving the data from the server and without having the server access the entire file.

In the existing cloud storage system, the owner want to check the data integrity, he need to access the entire file so it's expensive to the cloud server. Also transmitting the file across a network may consume high bandwidth. It's further complicated for the owner of the data whose devices like Personnel Digital Assist and mobile phones. Because these devices can have only a limited amount of battery power, CPU power, storage capacity and communication bandwidth. Basically using cloud storage, the owner stores their files in the cloud. Owner can check over the data integrity by enabling a new role which is TPA [5] because it possesses experience capabilities that the customer does not.

Third Party Auditors can understand the threats and they know best practices to identify the threats. Also they have the resources to check for process adherence and service quality. The TPA will be able to verify over any threats in online storage services that are represented in the cloud server. Thus, the user who owns the data can rely on the TPA to verify the data in the cloud without involving with the procedure. The encryption idea is based on scrambling the information that only the one who have the secret key can expose it by decryption. The encryption concept will not be enough to ensure the data integrity over the cloud. Sometimes TPA may modify file and upload it in cloud again

4. Proving the data integrity in cloud storage

Juels and Kaliski [3] proposed a scheme called Proof of Retrievability (POR). Proof of retrievability means Verify the data stored by user at remote storage in the cloud is not modified by the cloud. POR for huge size of files named as sentinels. The main role of sentinels is cloud needs to access only a small portion of the file (F) instead of accessing entire file. Sravan and saxena[6] proposed a Schematic view of a proof of retrievability based on inserting random sentinels in the data file. Semantic view of POR is shown in Fig. 2.

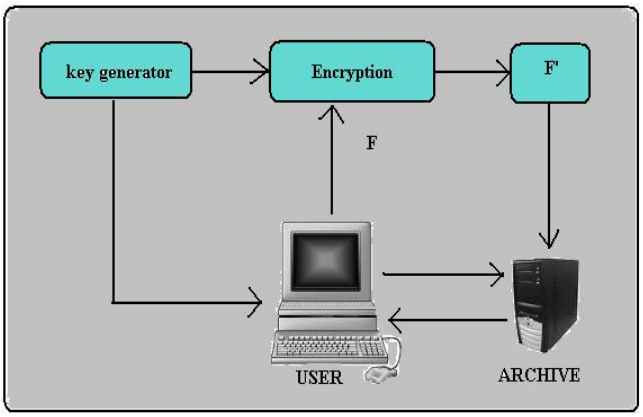

Fig.2 Schematic view of a POR

The above architecture describes that, user (cloud client) likes to store a file (F) in the cloud server (archive). Before storing the file to the cloud, owner needs to encrypt the file in order to prevent from the unauthorized access.

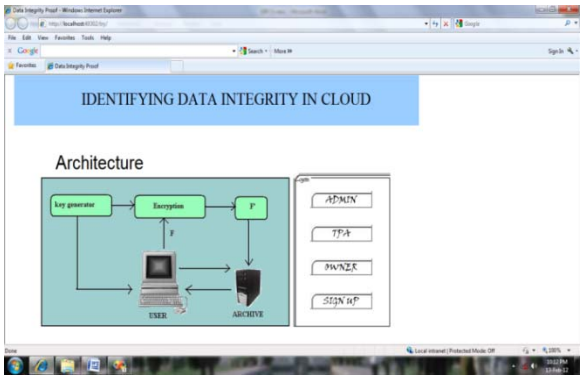

Fig. 3 Identifying data integrity in cloud

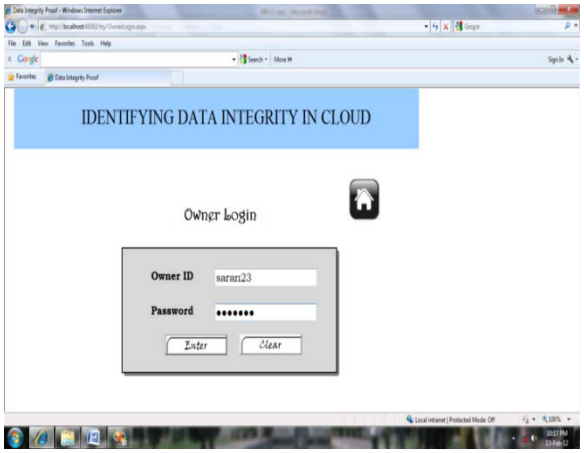

Fig. 4 Owner Login form

The current section discusses various aspects that should be considered to achieve data integrity. Company who is wishes to go for cloud storage service must be an authorized user and register themselves as a client. For every authorized user the system will generate a security key. Secret key is used while owner needs to login see fig. 3 and 4. The owner can get the secret key either through offline see fig.5 (a) and (b) or online.

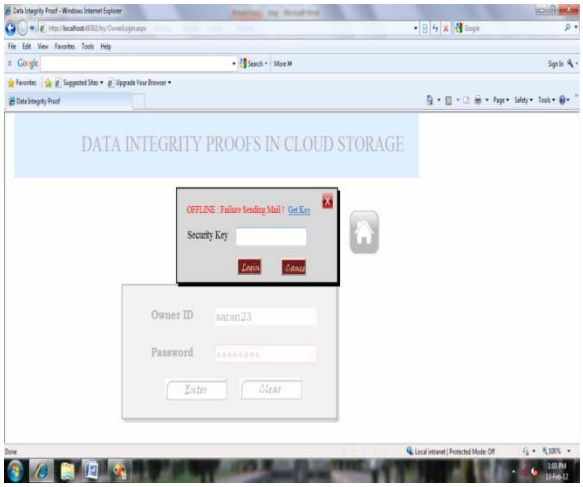

Fig. 5(a) Getting secret key while owner in offline

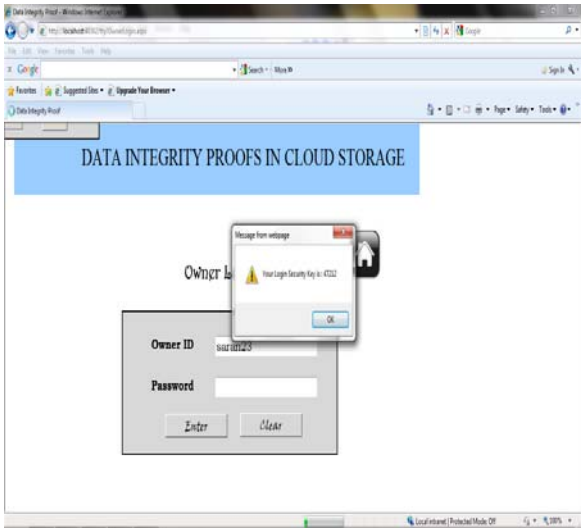

Fig. 5(b) Getting secret key

If the owner is in online, the secret key is sent to their mail as shown at the figure 6(a) and (b)

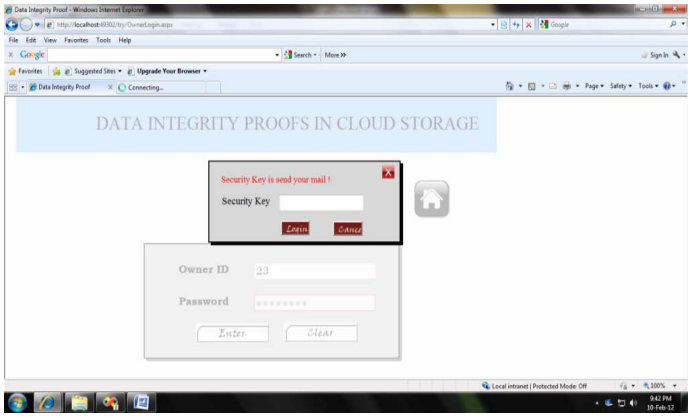

Fig. 6(a) Getting secret key while owner in online

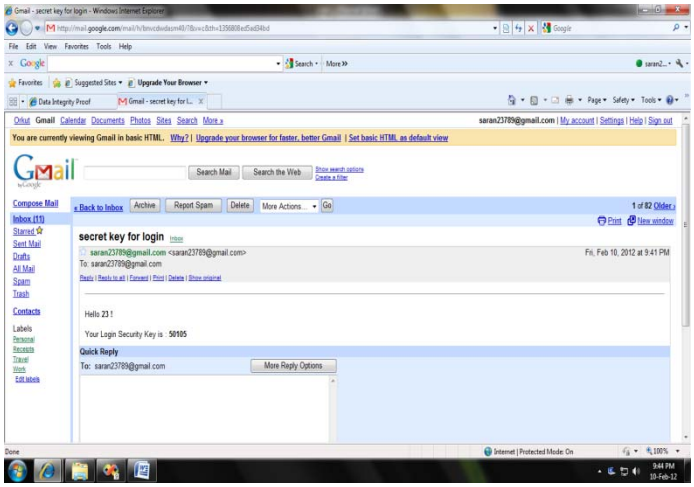

Fig. 6(b) Getting secret key through mail.

The proposed system ensures that unauthorized users are not permitted to login. The authorized client can upload the file into cloud. At the time of uploading the files into the cloud, the proposed system's key generator generates an encryption key and sends to the owner.

For every file which are uploaded in the cloud, TPA verifies it whether it is secured or not. This verification process can be done in two ways 1) direct verification and 2) download verification. Shown in figure 7 and figure 8.

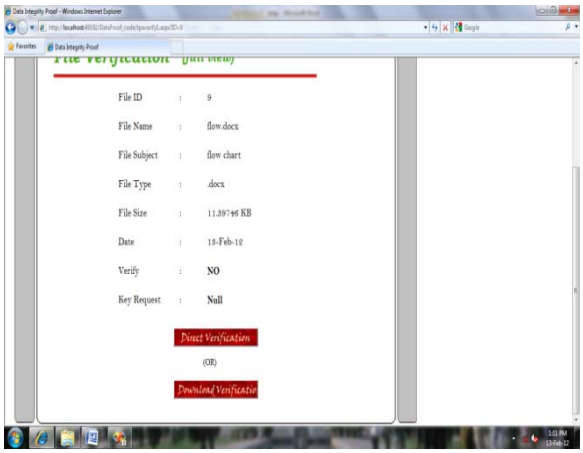

Fig. 7 File verification form

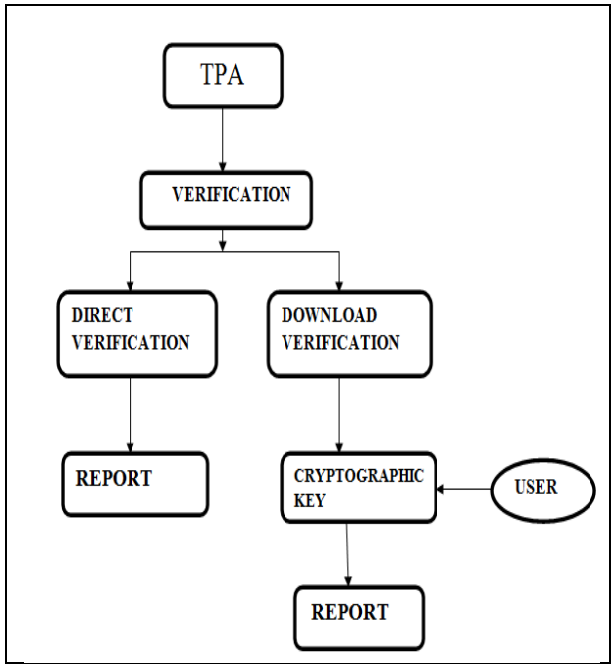

Fig. 8 Two types of file verification

In direct verification, TPA directly verifies the file without the need of cryptographic key. In the download verification, cryptographic key is necessary to download the file. For cryptographic key, TPA needs to send key request to the owner. Owner sent the key to the TPA to check the integrity of the file and sends report to the owner. Some times TPA may alter the file and upload the file again in the cloud. If the TPA modified the file means cloud sent alert to the owner. Through this process integrity of the file is identified and assured.

Algorithm for cloud storage

File is denoted as 'F'
Owner of data is represented as 'cc'
Cloud-server is denoted as 'cs'
Secret-key is represented as Skey and
Encryption-key as Ekey
begin
If uservalue==Skey then
Login:=true;
else
Report:=Invalid owner
End if
cs<--F'(F U Ekey);
end;

Algorithm for verification

TPA is used by 'cc' to verify the integrity
begin
if verifyproof=direct then
report:=direct access of file
else
Return {1,0}<--verifyproof(Ekey)
*/*outputs 1(TRUE) if the integrity of the file is verified as correct, otherwise 0 (FALSE).*/*

5. Conclusions

The next generation of cloud storage provides a new architecture to address the storage, management and analysis of fast-growing machine-generated data. This paper briefly explaining about the cloud storage, advantages along with its characteristics. The proposed system provides the proof of the data integrity and the owner can check the integrity of their data in efficient manner. If any modifications did by the TPA, cloud will immediately intimate to the owner of the file. So Security and data integrity is secured properly. And it reduces the access time at the cloud server and reduces the cost for retrieving the file and bandwidth consumption across the network.

References

[1] Paul Zimski, “Cloud computing faces security storm” in 2009.

[2] Jia xu and Ee-chien chang, “Towards efficient proofs of retrievibility in cloud storage”.

[3] A. Juels and B.S. Kaliski, Jr., “Pors: proofs of retrievability for large files,” in CCS ’07: Proceedings of the 14th ACM conference on Computer and communications security.

[4] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, and D. Song, “Provable data possession at untrusted stores,” in CCS ’07.

[5] Dalia Attas and Omar Batrafi “Efficient integrity checking technique for securing client data in cloud computing” in IJECS-IJENS, 2011.

[6] R. Sravan kumar and Saxena ,”Data integrity proofs in cloud storage” in IEEE 2011.