Data Integrity Proof (DIP) in Cloud Storage

Pratik Bhujbal[1] Anupam Pahnale[2] Parve Gorakh[3] Prof. Mohit Dighe[4]
[4]Assistant Professor
[1,2,3,4]Department of Computer Engineering
[1,2,3,4]SCSCOE, Rahuri Factory, India

Abstract— Nowadays, Data is rising over internet in terabytes and Exabyte. So, there is a need of storing these data which has been fulfilled by cloud computing. Though the service of cloud appear to be efficient and cost effective. Yet there are some challenges which are faced in cloud computing such as Data security and Authentication. In cloud Storage the data of Owner is stored in cloud where the cloud servers are remotely located the owner of the data does not have any direct control over the data. If the data over cloud is modified by the cloud, Third Party Auditor (TPA) or any other person there is no precision such that the owner of the data gets the information about the modification of the data. TPA is a Third Party Auditor who has Experience in checking the integrity of data. TPA verifies the files stored over the cloud if they are modified or not. Our scheme provides the solution to this problem such that if there is any modification in the data the owner will get information about the change in the data. Our scheme only provides the information about the change in data it does not keep data intact or secure from modification over cloud.

Key words: Third Party Auditor (TPA), Cloud Storage, Cryptography, Data integrity

I. INTRODUCTION

The usage of computers, mobile gadgets and social networking sites is now part of common mans day to day life. Sharing of information, photos, video and audio files have enabled user to communicate and utilize virtual storage space in the Internet without worrying to buy physical storage locally. All these data need to be stored somewhere in Internet and Cloud happens to be a default choice.

In order to reap the operational and financial benefits of Cloud, the enterprises are also storing data with third parties in Cloud. It is challenging for small and medium company to keep updating hardware according to increasing data [1]. Cloud service provides flexibility to use storage services on demand according to ever changing requirement of enterprise. Storage as service is popular model, where data storage is outsourced by enterprise to third party service provider (Cloud Provider) [2] who charges as per the usage of storage facility. On the other side of spectrum, there is increasing trend these days, in the form of ubiquitous presence of mobile devices and the wide variety of functions for which they are used. Most of these functions are data generating (like photography, video shooting etc

A. Data Integrity in Cloud Storage

Integrity, in terms of Network and data security, is assurance that information could only be accessed and modified by those authorized for it. Measures are taken to ensure Data integrity includes controlling physical environment of network terminals and the servers, restricting access of data, and maintaining strict authentication practice. Data integrity can also threatened by environmental hazards, such as heat, dust, and electrical surges. Data Integrity is most important of all security issues and privacy in cloud data storages because it not only ensures completeness and correctness of data but also ensure that data is consistent, correct, correct and of high quality.

II. LITERATURE REVIEW

In Literature survey the study of cloud services and study of data integrity proof in cloud storage. Many solutions have been provided to focus on resolving the issues of integrity. Juels and Kaliski[1] proposed a model Proofs of Retrievability(POR) was one of the first most important attempts to formulize the notion guaranteed remotely and reliable integrity of the data without the retrieving of data file. It is basically a data encryption mechanism which detects data corruptions and retrieve the complete the data without any damage. Shacham and Waters[2]gave a new model for POR enabling verifiability of unlimited number of queries by user with reduced overhead. Later Bowels and Juels[3] gave a theoretical model for the implementing of POR, but all these mechanisms proposed were weak from the security point because they all work for single server. Therefore Bowels [4] in their further work gave a HAIL protocol extending the POR mechanism for multiple servers. Priya Metri and Geeta Sarote[5] proposed a threat model to overcome the threat of integrity and provide data privacy in the cloud storage. It uses TPA(Third Party Auditor) and digital signature mechanism for the purpose of reliable data retrievable. The TPA being used notifies any unauthorized access attempting to make changes, avoiding the changes in data and maintaining the originality of data. Atienies and Burns[6] gave Provable Data Possession(PDP) mechanism which verifies the integrity of data being outsourced, detecting all kind of errors occurring in data but doesn't guarantee complete data retrievable. In their later work Atienies and Pietro[7] proposed a scheme which overcome all problems in PDP, but the main and basic problem on both proposed system didn't overcome was they work on single server. Therefore, later Curtmola[8].

III. DATA INTEGRITY PROOF SCHEMES

A. Overview of Data Integrity Schemes

As the word suggests itself data integrity means completeness or wholeness and it is basic requirement of information technology [7]. Data integrity refers to maintaining and assuring the accuracy and consistency of data over its entire life-cycle [8]. Data corruption is a form of data loss and data integrity is opposite of data corruption [8]. Data integrity ensures the data is the same as it was when it was originally recorded.

B. Proof of Retrievability

In the Proof of Retrivability (POR) scheme this scheme using keyed hash function is the simplest scheme than any other scheme for proof of retrivability of data files [5]. In this scheme the data file is stored in the cloud storage but before storing it in the cloud storage that file is pre-processed and Encryption cryptographic hash is computed [5]. After calculating hash value the file is stored in the cloud storage or Data Center [5]. The Encryption cryptographic key which is used to calculate hash value is then released to cloud storage and value calculated by the cloud storage are compare with each other [5]. From that comparison final conclusion is considered [5]. The main advantage of this scheme is simple to implementation. Limitation of this scheme is, it is computational burdensome or difficult for devices like Laptops, mobile phones, PDAs etc. [5]. Next scheme for proof retrivability is using position of bits or sentinels [3]. This concept is proposed by Ari Juels and Burton S. Kaliski Jr [3]. Sentinels are the special blocks which are used in this scheme to verify the integrity. Sentinels are embedded in the data blocks randomly during setup phase by the verifier in the setup phase [3]. The integrity of the data file is calculate by challenge and response. The verifier or TPA throws challenge to the cloud storage by specifying the position of the collection of the sentinels and the cloud storage has to return the associated sentinels or TPA values to the verifier[3][5]. If the file stored by the client is modified then the associated sentinel's values also get changed and the cloud will return wrong values to verifier.

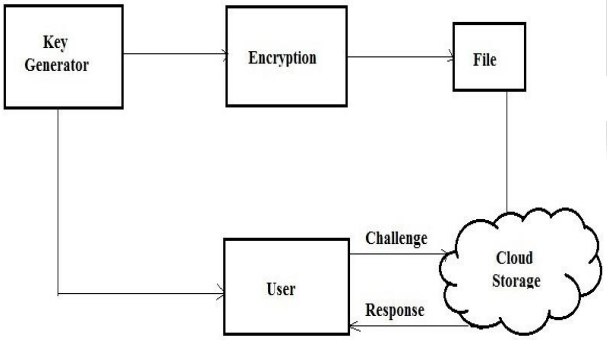

Fig. 1: A Diagram of a proof of retrievability

Insert random sentinels in data files F

From this integrity of the file is checked [3][5]. Limitation of this scheme is that this scheme involves encryption of file so this is computationally cumbersome for the small devices like mobile phones, PDA etc. [5].

Also Sravan Kumar R. and Ashutosh Saxena present this scheme [5] which involves selection of random bits per blocks of data due to this computational overhead of the client is reduced. File is processed by the verifier before storing it in the cloud storage [5]. After that verifier attach some metadata to the file [5].This meta data is used at the time of verification of the integrity of the file [5]. The limitation of this scheme is this scheme applies only static data [5].

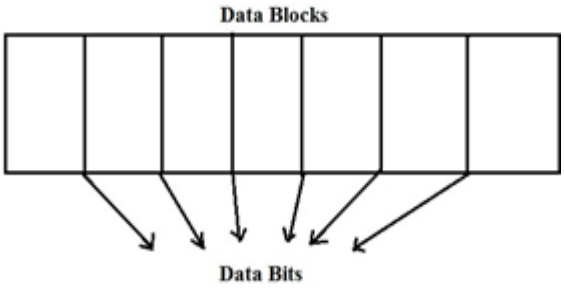

Fig. 2: File is divided into number of Data Blocks

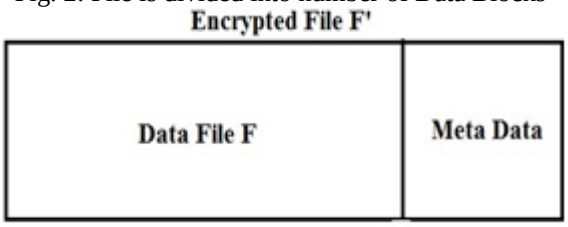

Fig. 3: File is appending the Meta Data and made Encryption file F'

1) Setup Phase

Let the verifier V wishes to the store the file F with the archive. Let this file F consist of n file blocks. We initially preprocess the file and create metadata to be appended to the file. Let each of the n data blocks have m bits in them. A typical data file F which the client wishes to store in the cloud is shown in Figure 2.

2) Metadata Generation Phase

Let g be a function defined as follows

$g(i, j) \rightarrow \{1..m\}, i \in \{1..n\}, j \in \{1..k\}$ .....(1)

Where k is the number of bits per data block which we wish to read as meta data. The function g generates for each data block a set of k bit positions within the m bits that are in the data block. Hence g(i, j) gives the jth bit in the ith data block. The value of k is in the choice of the verifier and is a secret known only to him. Therefore for each data block we get a set of k bits and in total for all the n blocks we get n*k bits. Let mi represent the k bits of meta data for the ith block. Fig. 3 shows a data block of the file F with random bits selected using the function g.

3) Verification Phase

Let the verifier or TPA V wants to verifying the integrity of the file F. It throws challenge to archive and asks it to respond. The challenge and response are compared and the verifiers accept or reject the integrity proof. Suppose verifier wishes to check integrity of nth block. The verifier challenge cloud storage server by specify the block number i and bit number j generated by using function g which only verifier knows. The verifier also specifies position at which the Meta data correspond the block i is append. This Meta data will be a k-bit number. Hence cloud storage server is required to send k + 1 bits for verification process by the client. The Meta data sent by cloud is decrypted by using number i and corresponding bit in this decrypted Meta data is comparing with bit that is sent by the cloud. Any not match between two would mean a loss of the integrity of the client data at the cloud storage.

IV. PERFORMANCE AND ANALYSIS

First we Analyze our system for Owner Registration response for registration by system is pretty fast our anlysis is shown following table.

Action Performed	Response	Remark	Result
Owner Registration	Owner Record Saved	Fast	Registration Successful

Table 1. Analysis table for Owner Registration

The we analyze our system for login of owner where often using Login-ID and Password owner receive One Time Password (OTP) on send to Mail-Id which has to be given for login of user. Our Analysis is shown in table.

Action Performed	Response	Remark	Result
Owner Login	OTP send on Owner Mail	Fast	Login Successful

Table 2. Analysis table for Owner Login

Owner has perform some other operations such as File Upload . and After receiving the request for download file of TPA either Owner can accept or reject the request.

Action Performed	Response	Remark	Result
File Upload	Cryptography key Send on Mail	Medium	File upload Successfully
File Download	Downloaded File	Medium	Successful
TPA Request Accepted	Cryptography key send to TPA	Fast	Successful

Table 3. Analysis table for Owner Operations

Third Party Auditor (TPA) performs the operation of file verification where TPA has precision to either verify file diectly or by downloading the file. The analysis TPA shown Following table.

Action Performed	Response	Remark	Result
Direct Verification	Directly Verify the File	Does not support all types of the file.	Failure for pdf, audio and video files
Downloaded Verification	Cryptography key and request send to owner	Slow process as requires permission of owner	Successful

Table 4. Analysis table for TPA Operations

Admin has Database of file modification if Admin sense any modification in any of file in database Admin sends the warning of modification of file to the owner.

Action Performed	Response	Remark	Result
If File is Modify or Not	Send Warning to owner File is Modify or Not	Medium	successful

Table 5. Analysis table for Admin Operations

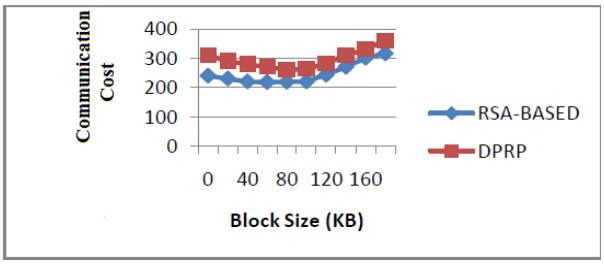

Fig. 4: shows communication cost over block size

As can be seen in fig. 2, it is evident that the proposed RSA based scheme communication cost of DPRP scheme is more when compared with the proposed RSA based scheme. RSA based approach is yielding more performance.

V. Conclusion

In this paper, we have defined next generation for cloud storage which provides a scheme to address storage, management and the analysis of the rapidly growing machine Generated information. This paper explains about cloud storage, advantages along with its characteristics. Our scheme is to reduce computational data and the storage Overhead of the cloud storage server. We also have minimized the size of proof of data integrity so that to reduce network bandwidth consumption. At client we only store the two functions, bit generator function g, and function h which can be used for encrypting data. Hence storage at client is very small as compared to other schemes that were developed. In our scheme encryption task is very limited to only fraction of whole data thus saving computational time of client. Many of schemes earlier require the archive to perform processes that need lot of computation power to generate proof of the data integrity. But in our scheme archive just need to fetch and send few bits of data to client. And also evaluate performance of cloud storage performance.

References

[1] Daliya Attas, Omar Batrafi, Efficient integrity checking technique for securing client data in cloud computing, IJECS Vol:11 No:05.,2011.
[2] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, and D. Song, Provable data possession at untrusted stores, New York, NY, USA: ACM, 2007, pp. 5986095.
[3] A. Juels and B.S. Kaliski Jr., Pors: Proofs of Retrievability for Large Files, Proc. 14th ACM Conf. Computer and Comm. Security, pp. 584-597, 2007.
[4] Amazon.com. Amazon simple storage service (Amazon S3), 2007.
[5] Sravan Kumar R, Ashutosh Saxena, Data Integrity Proofs in Cloud Storage,2011
[6] Satyakshma Rawat, Richa Chowdhary, Dr. Abhay Bansal, Data integrity of cloud data storage (CDSs) in cloud ijarcsse Vol. 3, Issue 3, March 2013.
[7] Saranya Eswaran, Dr. Sunitha Abburu Identifying Data integrity in cloud storage, IJCSI Vol. 9, Issue 2, No 1, March 2012.
[8] E. Mykletun, M. Narasimha, and G. Tsudik, Authentication and integrity in outsourced databases, Trans. Storage, vol. 2, no. 2, pp. 107138, 2006.

