International Research Journal of Computer Science (IRJCS)
Volume 1 Issue 1 (September 2014)
ISSN: Applied
www.irjcs.com

A Secure & Efficient Data Integrity Model to establish trust in cloud computing using TPA

Mr.Mahesh S.Giri
Department of Computer Science & Engineering
Technocrats Institute of Technology
Bhopal, India

Dr. Setu Chaturvedi
Department of Computer Science & Engineering
Technocrats Institute of Technology
Bhopal, India

Abstract- ***Cloud computing comes into focus only when we think about what IT always needs: A way to increase capacity or add capabilities on the fly without investing in new infrastructure, licensing new software. Besides of this advantage there is one major problem we have to face while keeping our private data in cloud, Assurance of our data will remain unaffected by external as well as internal threat and security are major concerns as there are chances for CSP to behave unfaithfully with users regarding the status of their outsourced data. Client cannot physically access the data from the cloud server directly, without client's knowledge, cloud provider can modify or delete data which are either unused by client from a long a time or takes large memory space. Hence, there is a requirement of checking the data periodically for its integrity, checking data for correction is called data integrity. To overcome data integrity problem, the user of the data must be able to use the assistance of a Third Party Auditor (TPA). The TPA has an experience that and check the integrity that is not possible for small users to check. The user can authorized the integrity checking responsibility to the TPA, in such a way that the TPA will not be able to manipulate with the client data with one way or another .In this paper we will first provide an introduction about the cloud computing and data integrity. We will introduce a model for the integrity checking over the cloud computing with the support of the TPA and further describe about the future work in this area.***

Keywords- Cloud Computing, data integrity, Third Party Auditor,

I .Introduction

Cloud Computing is the next generation technology and emerging as future of computing World, due to its advantages-on demand service, location rapid resource elasticity ,independent resource pooling, and pay and use based policy. It is derived from Grid

computing but still it make its own unique identity in IT industry and provides three service model SaaS, PaaS and IaaS In this paper section 1.1 explains about what cloud computing is and what it deals with, section 1.2 deals with challenges in cloud commuting section 1.3 discuss about data integrity and objectives of integrity . The sections of the paper is organized as follows: section two describes the proving schemes currently imposed to ensure data integrity, section three focuses on our proposed mechanism for data integrity, section four describes the expected outcomes and the paper is concluded finally in section five.

I. Introduction to Cloud Computing & Data Integrity

1.1 Cloud computing

According to Hewitt, C.[1] cloud computing is defined 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 growing need of Technology in every field has lead to the evolution of cloud computing for highly efficient usage of IT resources. Cloud computing offers different types of services, including Software as a Service (SAAS), Platform as a Service (PAAS) and Infrastructure as a Service (IAAS). Cloud Storage is an important service of cloud computing, which allows data owners to move their data remotely. More and more data owners start choosing to host their data in the cloud.

1.2 Challenges in cloud computing

Every day, a fresh news item, latest publication ,blog entry, highlights the cloud computing's security risks and threats. In each technology there are some security issues that affect the usage and the behavior below some of these concerns in the cloud: [2]

- Access: When there is an unauthorized access to the data, the ability of altering on the client data arise.
- Availability: The data must be available all the time for the clients without having problems that affect the storage and lead to the client data lose.
- Network Load: The over load capacity on the cloud may drop the system out according to the high amount of data between the computers and the servers.

- Integrity: The data correctness, legality and security is the most fields that influence on the cloud and have major lay on the service provider.
- Data Location: The client does not know the actual place that the data saved or centered in because it distributed over many places that led to confusion.

One of the important concerns in the cloud computing that need to be addressed is to assure the customer of the integrity ,accordingly in the next section we will discuss about data integrity and its objectives

1.3 Data Integrity:

Integrity, in terms of data and network security, is the assurance that data can only be accessed or modified by those authorized to do so, in simple word it is process of verifying data. Data Integrity is very important among the other cloud storage issues. Because data integrity ensured that data is of high quality, correct, unmodified . After storing data to the cloud, user depends on the cloud to provide more reliable services to them and hopes that their data and applications are in secured manner. But that hope may fail some times (i.e.) the user's data may be altered or deleted. Sometimes, the cloud service providers may be dishonest and they may discard the data which has not been accessed or rarely accessed to save the storage space or keep fewer replicas than promised [3]. Moreover, the cloud service providers may choose to hide data loss and claim that the data are still correctly stored in the Cloud. As a result, data owners need to be convinced that their data are correctly stored in the Cloud. So, one of the biggest concerns with cloud data storage is that of data integrity verification at untrusted servers. In order to solve the problem of data integrity checking, many researchers have proposed different systems and security models. In this paper we assume that the data integrity in the cloud is well verified using the TPA, because it possesses experience capabilities that the customer does not.

II. Current Data Integrity Proving Schemes and challenges

In Cloud computing the issue of data integrity is still carried out by many researchers they are still working with many different techniques .In this section we try to focus on few such techniques .we provide survey on the two different techniques of data integrity and there challenges. The basic schemes for data integrity in cloud are Provable Data Possession (PDP) and Proof of Retrievability (PoR)

2.1 Provable Data Possession (PDP):

2.1.1 Definition:

In simple words Provable Data possession (PDP) is a technique for validating data integrity over remote servers. In PDP A client that has stored data at an unfaithful server can verify that the server possesses the original data without retrieving it. Ateniese et al. [5] are the first to consider public audit ability in their defined "provable data possession" model for ensuring possession of files on untrusted storages.

2.1.2 Working Principal:

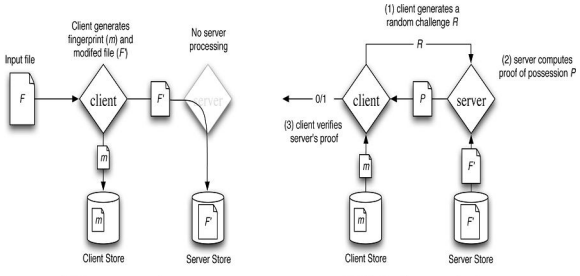
The working principal of PDP is as shown in figure 1.It works in two stages. Set up stage and challenge stage

Set up stage:

1. The client generate pair of matching keys public & secrete key by using probabilistic key generation algorithm
2. The client then sends the public key and the file to the server for storage and deletes the file from its local storage.

Challenge stage:

1. The client requests from the server a proof of possession for a subset of the blocks in the file.
2. The client checks the validity of the proof.



(a) Pre-process and store (b) Verify file possession by server

Figure 1 Protocol for Provable Data Possession [5]

2.1.3 Related Work and Challenges in PDP:

Ateniese et al model do not support dynamic data operations which are very important in cloud computing. It also suffer from design and security issues. . In their subsequent work [5], Ateniese et al. proposed a dynamic version of the prior PDP scheme problems by using symmetric key cryptography. This is lightweight than their previous model and allows for block updates, appends and deletions to the stored file However, their model targeted only on single server scenario and does not address small data corruptions, leaving the two most important issue unsolved i.e distributed scenario and data error recovery .
Curtmola et al. [6] extended the PDP scheme to cover multiple replicas without encoding each replica separately. This model is aimed to ensure data possession of multiple replicas across the distributed storage system. Even though this model provides guarantee that multiple copies of data are actually maintained on distributed server,it need third parties to conduct the audits which are again not the better option. .
In [4], Wang et al. consider the proposed challenge-response protocol which determine both the data correctness and locate possible errors. This challenge response protocol is almost used in every scheme with little modifications and they are having drawbacks of having large token size which unnecessary puts burden on clients.

2.2 Proof of Retrievability (PoR):

2.2.1 Definition:
Proofs of Retrievability (POR) is a cryptographic method for remotely verifying the integrity of files stored in the cloud, without keeping a copy of the user's original files in local storage. In a this scheme, a user backups his data file together with some authentication data to a potentially dishonest cloud storage server. Later, if user want to periodically and remotely verify the integrity of his data stored with CSP using the authentication data, without retrieving back the data file from cloud(6)

2.2.2 Working Principal

A POR works on two phase first is setup phase and another is sequence of verification phases.

Setup phase
In the setup phase, user preprocesses his data file using his private key to generate some authentication code. Then he sends the data file together with authentication code to the cloud storage server, and removes them from his local disk. Consequently, in the end of setup phaseuser has his private key in her local disk, and CSP has both the data file and the corresponding authentication code.

Sequence of verification phases:

In each sequence of verification phase, user generates a random challenge query and CSP is supposed to produce a short response or proof upon the received challenge query, based on user's data file and the corresponding authentication information. In the end of a verification phase, user will verify CSP's response using his private key and decide to accept or reject this response coming from CSP

2.2.3 Related Work and Challenges in POR.

Proofs of Retrievability (POR) model proposed by Juels and Kaliski [7] is among the first few attempts to formulize the notion of remotely and reliably verifying the data integrity without retrieving the data file. POR can be formalize By randomly embedding "sentinel"
blocks into the outsourcing file and hiding these "sentinel" blocks' position by encryption, their scheme can detect static data corruption effectively. However, their scheme cannot support any data update, and the number of queries a client can perform is fixed.

Shacham and Waters [8] design an improved PoR scheme with full proofs of security in the security model defined in [8] called as Compact PoR with rigorous security proofs. Based on the BLS signature, they aggregate the proofs into a small value and their scheme can support public verifications. However, using their scheme in dynamic scenario is impractical and insecure

III: Proposed Mechanism For Data Integrity

Based on the PDP and PoR schemes that we have discussed in this paper we come to following conclusion

1. PDP scheme easily support dynamic operation where PoR scheme is not support dynamic operation ,
2. PDP Scheme does not include error-correcting code whereas error correction can be done in POR more effectively .

Based on this conclusion in this paper, we propose a scheme which will enable the users to verify intactness of their data which is hosted on third party data centers. We are using Proof of Retrievibilty (POR) and Provable Data Possession (PDP) schemes as a reference to our work and based on these schemes we are proposing an approach for securely storing of data in clouds.

3.1 System Model:

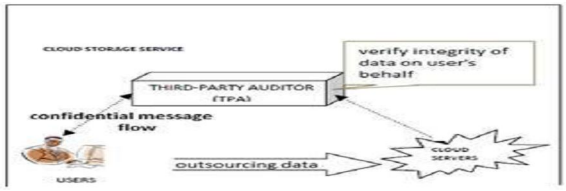


Figure 2. Architecture of cloud data storage service

In Figure 2, the proposed architecture is introduced to provide data integrity in could system which has three main roles as below

1. User: Users are the data owners who have the large amount of data to be stored in the cloud and relies on cloud storage server for data computation and maintenance, can be either individual consumers or organizations.
2. Cloud Storage Server: A cloud storage server (CSS) is an entity that is managed by cloud service provider (CSP). It provides space to the user for data storage and computation.
3. Third Party Auditor (TPA): An TPA is an entity who has expertise and capabilities that user may not have, is trusted to access or expose the risk of cloud storage services on behalf of client upon request.

In cloud data storage, a user stores his data on set of cloud servers in in this overall process CSP play as role of mediator between user and cloud server. As user no longer possess his data locally,he is totally depend on CSP that his data will remain as it is on cloud server.So in this case data integrity of user data plays very important role to assure user that his data are being correctly stored and maintained. User can check over the data integrity by enabling a new role which is TPA because in this proposed mechanism we are considering user with small device on which data integrity operation cannot be performed and user with small devices cannot monitor their data.The TPA has an access to the cloud provider environment and understands the service level agreements (SLA) that is between the customer and the provider. By this way the TPA is reliable, and independent [8] .TPA is expertise in checking the data integrity of user data they are having all the resources and computation power as well as bandwidth for monitoring user's data. 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. In our model, we assume that the TPA and cloud server communicate with each other to provide integrity and the user will authenticate them each time.

3.2 Implementation

In our proposed system we are implemented a model which establishes the trust in cloud computing by allowing the TPA to verify the correctness over the cloud data. In this proposed model we need to care that the cloud server will not alter or modify user's data.To achieve this goal we deal with the problem of implementing a protocol for obtaining a proof of data possession in the cloud it is also called as Proof of retrievability (POR)

There are three actors in our model to satisfy the data integrity concept:

1. User will do the registration with cloud and TPA by using his mail id as username.
2. The TPA verifies over the cloud server part to check if the cloud server was manipulating in the user data or not.
3. CSP will authenticate and verify TPA whenever client connects to cloud.

Now we will try to focus on each of this entity in details

User: User first register with the TPA by mentioning the url of CSP.User then register with CSP by mentioning the url of TPA. Email ID of user will be userid at TPA and public cloud. And same email id should be used at TPA and cloud.
CSP: The CSP will store user's file. Whenever the client register with cloud CSP will authenticate the user as well as the TPA as user is mentioning the url of TPA while registering.CSP will come to know that particular user is assign with particular TPA
TPA: Whenever user have file to upload on cloud it will redirect to TPA first.TPA will encrypted it.TPA will apply data integrity constraints on encrypted file and store signature/integrity data on TPA. After finishing the Data integrity check , the TPA will inform the user to his register mail id whenever the file is modified or alter on cloud storage.

3.2.1Data Integrity check at TPA

In our proposed model whenever user wish to upload a file to cloud storage it will redirect to TPA for encryption. As we are considering that user of the data may be a small device, like a PDA or a mobile phone as they are not able to perform any operation on their file. In our system TPA is responsible for performing data integrity check on the behalf of user. In our proposed model we do not encrypt the whole data. We only encrypt only few bits of data per data block thus reducing the computational overhead on the TPA. The TPA storage overhead is also minimized as it does not store any data with it. It will store only encryption key. Hence our scheme should minimize the local computation at the TPA as well as the bandwidth consumed at the TPA. In the next section we will focus on how the integrity checksum is done at TPA.

Setup Phase

The main idea is to compute some metadata on input file before outsourcing it to cloud storage. This metadata is nothing but some checksum sometimes called as token. To calculate this token many techniques used. We are using AES to calculate these tokens. Tokens are not big values; they are very small in size. So it will not put much burden on TPA to store these token. Input file is divided into some t blocks and tokens are calculated for each block. Later these tokens are stored at client side and the input file is outsourced to cloud storage. But before sending these tokens files are modified to identify the corresponding token and block pairs. This phase we called as Setup Phase as we are setting the environment to ensure correctness or intactness of our data. Setup Phase is explained in the Fig.3. As shown in Fig.4 the sample scenario of these token calculations. File F is divided in four parts, (We have divided the file in only 4 parts for the sake of simplicity.) namely F1, F2, F3, F4.

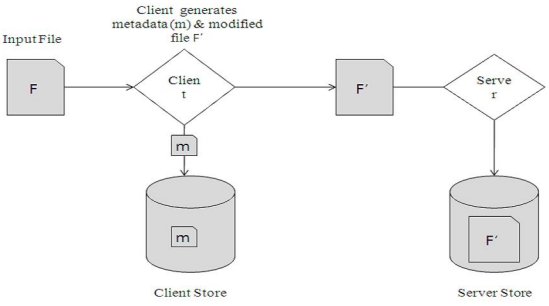


Figure 3. Setup Phase

Now we have calculated Hash on each of the block of file F. Now these Hash values are nothing but tokens for particular blocks of file. These pre-computed tokens are very small in size
and they are stored at the owner's database. As we know that tokens are of small size, it will not keep any memory burden on TPA.

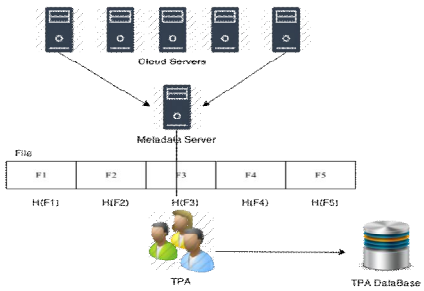


Figure 4. Example of Setup Phase

Whenever TPA wants to check the correctness of the data, he need not search whole data, instead he can check for specific block of data. So TPA will just use the locally stored token and will generate a random nonce. Finally he will attach nonce with the token to avoid the replay attacks from servers.

Verification Phase

To check the intactness of the users data; TPA can challenge the server. This process is called as Verification Phase. In this phase TPA generates random challenge on required block of
file. This Challenge is given to server. In response to this challenge server computes the proof of possession of data and sends back to TPA. This reply from server is matched with token stored at owner side. If owner found match then the data is intact. The process of Verification Phase is shown in Fig.5. This process is also called as Challenge Response Protocol Phase.

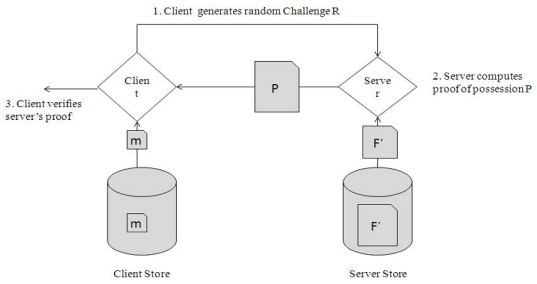


Figure 5. Verification Phase

The main idea is that, before outsourcing data file to cloud the user will ask TPA to perform data integration, in our data integrity protocol the verifier needs to store only a single cryptographic key - irrespective of the size of the data file.in this model we are only checking integrity of the data. This model does provide any prevention technique for altering data at cloud server. The proposed approach also supports such dynamic data operation in cloud. Suppose for example, if new data is added in the cloud or data is modified by the user of the data then, in that case before updating the data to cloud it will redirect to TPA which will then perform dynamic operation on that data, TPA will compute new tokens for the data. So it will help to accommodate dynamic behavior.

IV: Outcomes

With propose system, we can achieve an effective and flexible distributed scheme with explicit dynamic data support to ensure the correctness of users' data in the cloud. To ensure the security and dependability for cloud data storage in our proposed model, we aim to design efficient mechanisms for dynamic data verification and operation and achieve the following goals:

- **Data Integrity:** It ensure users that their data are safe and unaltered.
- **Error Recovery:** to effectively locate the dishonest server when data corruption has been detected.
- **Dynamic operation:** to maintain the same level of storage correctness assurance even if users modify, delete or append their data files in the cloud.
- **Dependability:** to enhance data availability
- **Lightweight:** to enable users to perform storage correctness checks with minimum overhead.

V: Conclusion

There are so many techniques available in the literature, out of which we have analyze Provable Data Possession (PDP) and Proof of retrievability(POR), This paper facilitate the client in getting a proof of integrity of the data which He/She wishes to store in the cloud storage servers with bare minimum costs and efforts. The scheme used in this paper reduce the computational and storage overhead of the client as well as to minimize the computational overhead of the cloud storage server. This also minimized the size of the proof of data Integrity so as to reduce the network bandwidth consumption. We have used a challenge-response protocol to ensure the intactness of data. This approach causes minimum overhead and also minimizes the bandwidth use. As we know cloud computing is not just third party data warehouse, it must support dynamic data updates. Our framework fully supports dynamic operations on data block which are very efficient. So the intactness of data is verified and along with that it provides protection against server colluding attacks which are more difficult to deal with. To summarize, the work described is an important step forward towards practical provable data possession techniques. We expect that the salient features of our scheme make it attractive for realistic applications.

REFERENCES:

1. Hewitt, C. (2008) ”ORGs for scalable,robust,privacy friendly client Cloud Computing Environment”.
2. Chandran S. and Angepat M., “Cloud Computing: Analyzing the risks involved in cloud computing environments,” in Proceedings of Natural Sciences and Engineering, Sweden, pp. 2-4, 2010.
3. Kan Yang · Xiaohua Jia, “Data storage auditing service in cloud computing: challenges, methods and opportunities”, DOI 10.1007/s11280-011-0138-0.
4. Kan Yang · Xiaohua Jia, “Data storage auditing service in cloud computing: challenges, methods and opportunities”, DOI 10.1007/s11280-011-0138-0.
5 G. Ateniese, R. D. Pietro, L. V. Mancini, and G. Tsudik, “Scalable and Efficient Provable Data Possession,” Proc. of SecureComm ’08, pp. 1– 10, 2008.
6 Chandran S. and Angepat M., “Cloud Computing: Analyzing the risks involved in cloud computing environments,” in Proceedings of Natural Sciences and Engineering, Sweden, pp. 2-4, 2010.
7 G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, and D. Song, “Provable Data Possession at Untrusted Stores,” Proc. Of CCS ’07, pp. 598–609, 2007.
8 Shah M., et al., “Auditing to keep online storage services honest” in Proceedings of HotOS’07, Berkeley, CA, USA, pp. 1-5, 2007.
9 R. Sravan kumar and Saxena ,”Data integrity proofs in cloud storage” in IEEE 2011.

