ISSN 2348 - 7968

# ENSURING DATA INTEGRITY IN CLOUD DATA STORAGE

**Gaurav Pachauri[1], Subhash Chand Gupta[2]**
**[1] Amity University Noida, India**

**[2]Amity University, Noida, India**

**Abstract :**Cloud Computing is the newest fast growing technology in the current era due to its easy manageable services. Cloud computing provides on demand network oriented services to centrally organised computing resources with great efficiency and minimal efforts. As cloud computing achieved through network access, there arises a lot of security concerns related to cloud computing. All data resides in centralized cloud data storage. This cannot be completely trustworthy .Cloud computing paradigm brings a lot of security challenges which are still to be resolved. The main concern is about the Integrity of data in cloud data storage. If a cloud service provider modify or delete our data from the storage due to some private problem then in that case how will we be able to verify that our data is modified or how will we be able to generate proofs that our data has been altered. These all are very serious issues related to Cloud computing .So in this paper we have discussed the current security mechanism to ensure the integrity of data in cloud storage. We have suggested a third party auditing mechanism to verify the integrity of data periodically without accessing the whole data. In prior mechanism to ensure the remote data integrity, there are also issues as some do not support public audibility and some do not support dynamic data operations. In this paper we have discussed all aspects of current mechanism as well as the third party involvement which resolves both public audibility and dynamic data operations.

**Keywords:** Auditor, Cloud Service Provider (CSP), data integrity, CDSs integrity.

## 1. Introduction

Cloud Computing is a computing system which is based on Internet. A lot of pooled resources will be available to a client on demand basis. Cloud computing provide software and services at a very low cost to its clients. Another benefit that comes in consideration is its simplicity. We do not require a lot of hardware to configure for using cloud computing services. In cloud computing data provided by Client will be stored centralized cloud data storage at remote servers. Basically cloud computing is a good mixture of early technologies for example Grid computing, Utility Computing and virtualization techniques. Some people still compare cloud computing with Grid computing but it is quite different with grid computing. We require only Internet access to use the services of cloud computing.

With all these ease provided by Cloud Computing, there arises issues related to security since Cloud computing is internet based facility. In this paper, our prime concern about the integrity of Data stored in cloud data storage.

Section 1 is about cloud computing, Section 2 is about the existing schemes on data integrity, section 3 is about our proposed scheme for verifying the correctness of data on cloud. In section 4 we have concluded the results.

ISSN 2348 - 7968

### 1.1 CLOUD COMPUTING

We can divide a cloud computing service according their functionality in these types as following.

**Infrastructure as a service(IaaS):** When a cloud service provider (CSP) provides infrastructure entities to clients such as server ,storage ,RAM, Virtualization techniques on rent ,then it is a infrastructure as a service of Cloud computing. An enterprise or client can use these facilities as a platform for building their applications.

**Platform as a Service (PaaS) :** In this cloud service provider provides their infrastructure for the deployment of users application. A user can easily use the environment provided by the cloud but the control on these facilities will be of cloud service provider. We can understand it with example of Google engine where a user can use the different APIs provided by Google for the development of applications. A user can use them to build applications, deploy them on Google servers and manage them.

**Software as a Service (SaaS):** In this a cloud service provider provides solution software running on cloud infrastructure .A user can access them through web browser and can change the settings according to their business needs .We can take the example of CRM provided by the sales force .

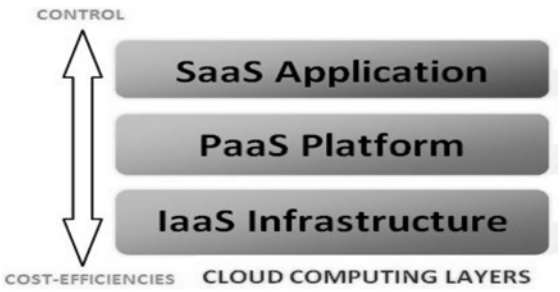



**Figure 1: Types of Services provided by Cloud**

### 1.2 Characteristics of cloud computing

Some of the important characteristics that a Cloud data Storages (CDSs) must possess are:

**Manageability:** The main work of Cloud data storage is to maintain the data provided by client since this data cannot be maintained on client's machine. So the data provided by client should be maintained by cloud system for this purpose cloud should be self-manageable in nature ,this self-managing of cloud system will reduce the burden of client.

**Availability:** A client will access their data regularly or frequently so this data should be available all the time to client easily. The data provided by Cloud should not be corrupted as this is the only backup for a client to get the data.

**Performance:** Since a cloud system uses internet services, and internet uses the TCP protocol for data transfer in the form of packets .the size of packets are small in nature so when we want to store large data it can be achieved through Cloud data storage which provides easy flow and storage for large data.

### 1.3 Features of Cloud Computing

There are five main features of cloud computing:

- On-demand self-service
- Broad network access
- Resource pooling
- Rapid elasticity
- Measured Service

**On-demand self-service:** Cloud computing services are on demand basis services. A client do not require any assistance from the cloud service providers to use these services.

**Broad network access:** Cloud computing resources are based on network services and they can be used in different manners with the help of different platforms like mobile phones, laptops etc.

**Resource pooling:** The providers provide resources in pooled manner to serve a lot of users simultaneously. These pooled resources will be available in the physical and virtual form dynamically assigned and reassigned according to

ISSN 2348 - 7968

consumer demand. In this sense, the customers do not have control or knowledge over the exact location of these resources.

**Rapid elasticity:** For a consumer, computing resources are elastic: they are scaled up to use whenever needed and scaled down to release whenever finished. To the consumer, resources provisioning often appears to be infinite and can be appropriated in any quantity at any time.

**Measured Service:** A cloud system has the capabilities to measure the utilization of particular customer. Cloud service provider has the mechanism to check the use of services for all customers and can be charged according to use . this mechanism will be transparent to both cloud and customer

## 2. Existing Systems and Glitches

Atinese et al [1] was first to provide a model for integrity check, he provide "Provable Data Possession (PDP)" model . In this model he used the public RSA algorithm for auditing outsourced data. In this manner he achieved the public audibility. But in this scheme they did not consider the dynamic data storage and when we move from static to dynamic data storage this scheme suffers design and security problems.

Later Atinese et al [2]presented a dynamic version of prior PDP scheme. However, the system imposes a priori bound on the number of queries and does not support fully dynamic data operations, i.e., it only allows very basic block operations with limited functionality, and block insertions cannot be supported. Wang et al. consider dynamic data storage in a distributed scenario, and the proposed challenge-response protocol can both determine the data correctness and locate possible errors. They only consider partial support for dynamic data operation.

This method describe[13] a "proof of retrievability" model, where spot-checking and error-correcting codes are used to ensure both "possession" and "retrievability" of data files on archive service systems. Specifically, some special blocks called "sentinels" are randomly embedded into the data file F for detection purpose, and F is further encrypted to protect the positions of these special blocks. the number of queries a client can perform is also a fixed priori, and the introduction of precompiled "sentinels" prevents the development of realizing dynamic data updates. In addition, public auditability is not supported in their scheme. Shacham and Waters, design[14] an improved PoR scheme with full proofs of security in the security model defined in previous method. They use publicly verifiable homomorphic authenticators built from BLS signatures, based on which the proofs can be aggregated into a small authenticator value, and public retrievability is achieved. Still, the authors only consider static data files. Erway et al. [15]were the first to explore constructions for dynamic provable data possession. They extend the PDP model to support provable updates to stored data files using rank-based authenticated skip lists. This scheme is essentially a fully dynamic version of the PDP solution. To support updates, especially for block insertion, they eliminate the index information in the "tag" computation in Ateniese's PDP model and employ authenticated skip list data structure to authenticate the tag information of challenged or updated blocks first before the verification procedure. However, the efficiency of their scheme remains unclear.

From the above studied schemes we can say that the existing schemes aim at providing integrity verification for different data storage systems, the problem of supporting both public auditability and data dynamics has not been fully addressed. Achieving a secure and efficient design to seamlessly integrate these two important

ISSN 2348 - 7968

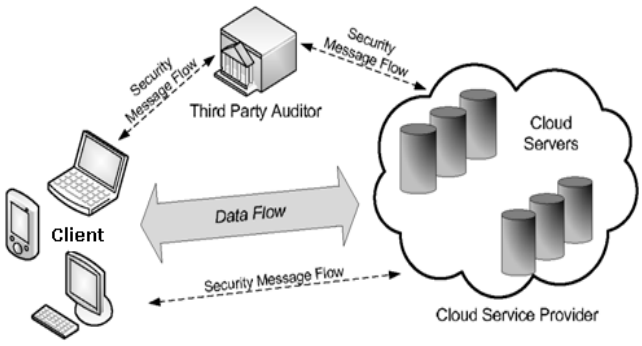
components for data storage service remains an open challenging task in Cloud Computing.

## 3. Proposed Scheme

In our proposed scheme we have divided the whole architecture of cloud in three basic entities, which are as following:

**Client:** An entity, which has large data files to be stored in the cloud and relies on the cloud for data maintenance and computation, can be either individual consumers or organizations.

**Cloud Storage Server (CSS):** An entity, which is managed by Cloud Service Provider (CSP), has significant storage space and computation resource to maintain the clients' data. (e.g. salesforce, gmail, amazon




The architecture of cloud data storage service

**Figure 2: Architecture of cloud data storage**

**Third Party Auditor:** An entity, which has expertise and capabilities that clients do not have, is trusted to assess and expose risk of cloud storage services on behalf of the clients upon request.

Once a client put his large data on remote servers he can be relieved with the burden of large data storage. But the client will not locally access their data so there should be some security measures to verify the integrity of data .For this purpose we use the third party auditor unit which will work on behalf of the client, the TPA have the right to verify the clients large data and asking for the proofs if any data is found modified. In this paper we have discussed the verifiability of the data that is public audibility. TPA will verify the files of a client with the help of public key. In this scheme we consider TPA as a unbiased entity and sever as an untrusted entity.

Public audibility in our system ensures that anyone who stored the data originally, not only client can perform block level operations on data which is stored. The design of the system is such that it supports dynamic data operations .A dynamic data operation allow a client to perform block level operations on data files .the design should be efficient enough to support both public audibility and dynamic data operations at the same level of integration. In our scheme TPA do not require to retrieve the whole file for the verification, this measure provides a level of security from TPA as TPA will not be able to see the files of user and hence cannot make any modifications to them.

A KeyGen()function is used to generate a public and private keys pair. A client will upload his data which will be encrypted at the same time of uploading. A hash will be generated for each file uploaded. For hash generation we use Merkel hash algorithm which works as follows: File is divided in to m blocks (X1, X2,X3….Xm) of n size as shown in figure 4.1The leaf nodes contains h(Xi) which is hash value of each data block starting frpm left to right. The parent hc, hd etc of each pair of leaf node is calculated using xor of two children. This is done repeatedly until root is created. The root hr is taken as the hash for the file. This hash is stored on cloud after signing root by clients' signature using a secret key which is secret between client and auditor.

The auditor views all files and chooses the file need to audit. Create a challenge to ask for the hash of the current file on cloud which also provides some auxiliary information to the blocks asked for verification. Cloud also sends the original root. Using this information auditor also generates new

ISSN 2348 - 7968

root (r1) of merkle hash tree corresponding to the current file on cloud. The root (r2) stored earlier by client on cloud is authenticated by auditor using secret key and mac verification. Proofgen()-The both r1 and r2 are matched by auditor to generate proof of integrity if both are same the data file is not corrupted else data file is corrupted and this proof is sent to the client .

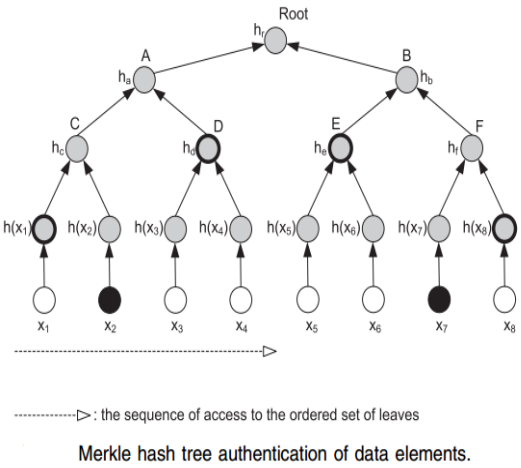



**Figure 3: Merkel Hash Tree authentication**

## 4. CONCLUSION

To secure cloud data storage it is a good practice to apply a third party auditing schemes. A client will leave the work of public audibility to third party auditor but sometimes they may also be unreliable and it is possible that they do not verify the client's data periodically. In this paper we have proposed simultaneous audibility with data dynamics for data stored in remote cloud computing storage. In this paper we tried to resolve both issues of public audibility and dynamicity of remote data. Efficiency is another major concern for this scheme as we kept in our mind the issue of efficiency also. For making the scheme more efficient we use the Merkel hash algorithm for tagging blocks for authentication. We can easily implement this scheme and can use for our purpose. We can connect this scheme with mobile gateway so that a TPA can immediately generate the reports to our mobile phone. Unlike most prior works, the new scheme further supports secure and efficient dynamic operations on data blocks stored in the cloud, including: data update, delete and append.

**ACKNOWLEDGEMENT**

I am highly grateful to my mentor Shubhash Chand Gupta who was continuously showing the right path to write this paper. I am highly obliged to Amity university library support which provided a lot of literature work to write this paper.

## References

**[1]** G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, and D. Song, "Provable Data Possession at untrusted Stores".

**[2]** Giuseppe Ateniese , Randal Burns, Reza Curtmola, Joseph Herring, Lea Kissner, Zachary Peterson, Dawn Song "Provable Data Possession at Untrusted Stores".

**[3]** Song, D. Shi, E., Fischer, I. Shankar, U., "Cloud Data Protection for the Masses", IEEE computer magazine, 2012.

**[4]** Q. Wang, C.Wang, Wenjing Lou, Jin Li, "Enabling Public Verifiability and Data Dynamics for Storage Security in Cloud Computing," IEEE transaction on parallel and distributed systems, VOL. 22, NO. 5, 2011.

**[5]** Cong Wang, S.M. Chow, Qian Wang,Kui Ren, and Wenjing Lou "Privacy-Preserving Public Auditing for Secure Cloud Storage", IEEE transaction,20 December 2011.

**[6]** Giuseppe Ateniese , Randal Burns, Reza Curtmola, Joseph Herring, Lea Kissner, Zachary Peterson, Dawn Song "Provable Data Possession at Untrusted Stores".

**[7]** Mehul A. Shah Ram Swaminathan Mary Baker Privacy-Preserving Audit and Extraction of Digital Contents.

**[8]** Alina Opre, Michael K. Reitery, Ke Yang, "Space-Efficient Block Storage Integrity"

**[9]** Danish Jamil et. Al., "Security issues in cloud computing and countermeasures", International Journal of engineering Science and Technology (IJEST).

**[10]** Alina Opre, Michael K. Reitery, Ke Yang, "Space-Efficient Block Storage Integrity".

**[11]** http://gadgets.ndtv.com/internet/news/gmail-is-down-again-in-india-with-502-error-366264?pfrom=home-editorpick

**[12]** Steve Fisher, " Security workbook ", FORCE.COM DEVELOPER'S LIBRARY online at http://developer.force.com/workbooks, 27 Oct 2012.

**[13]** 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.

**[14]** H. Shacham and B. Waters, "Compact Proofs of Retrievability," In Proceedings of Asiacrypt "08, Dec. 2008.

**[15]** C. Erway, A. Kupcu, C. Papamanthou, and R. Tamassia, "Dynamic Provable Data Possession".

**Gaurav Pachauri** is pursuing master of technology from Amity University Noida, India. Area of interests are cloud computing, Digital Image processing.

**Shubash Chand Gupta** is a scholar faculty in Computer science department, Amity University Noida, India.