International Journal of Science Engineering and Advance Technology

# Data Integrity in Multi Cloud Storage

**K. Venkat Rao, Avala Atchyuta Rao**
M.Tech Student, Dept. Of Cse, Assistant Professor, Dept. Of Cse,
Gokul Institute Of Technology And Sciences.

**Abstract**— Data integrity in cloud is became an acceptable challenge in the research community because of its vast application and usability in common people. This paper given an idea and ensure the secure and less loss in data during sharing with different user residing worldwide. We have introduced a new scheme called PDP (Provable Data Possession) scheme for distributed cloud storage to support the scalability of service and data migration. Here we have the existence of multiple cloud service providers to cooperatively store and maintain the clients' data.

We also introduce an another scheme called (CPDP) i.e. cooperative PDP which is based on homomorphic verifiable response and hash index hierarchy. We prove that the proposed scheme is minimizing the computational error as well as decrease the communication overhead as compare with existing methods.

**Index Terms**—PDP,CPDP, Multi cloud, Storage Security, computational error.

## 1.INTRODUCTION

Now a days , cloud storage service has become its low-cost, scalable, position-independent platform for clients' data. Since cloud computing environment is constructed based on open architectures and interfaces, it has the capability to incorporate multiple internal and/or external cloud services together to provide high interoperability. We call such a distributed cloud environment as a *multi-Cloud* (or *hybrid cloud*). Often, by using virtual infrastructure management (VIM) , a multi-cloud allows clients to easily access his/her resources remotely through interfaces such as Web services provided by Amazon EC2. There exist various tools and technologies for multicloud, such as Platform VM Orchestrator, VMware vSphere, and Ovirt. These tools help cloud providers construct a distributed cloud storage platform (DCSP) for managing clients' data. However, if such an important platform is vulnerable to security attacks, it would bring irretrievable losses to the clients. For example, the confidential data in an enterprise may be illegally accessed through a remote interface provided by a multi-cloud, or relevant data and archives may be lost or tampered with when they are stored into an uncertain storage pool outside the enterprise. Therefore, it is indispensable for cloud service providers (CSPs) to provide security techniques for managing their storage services.

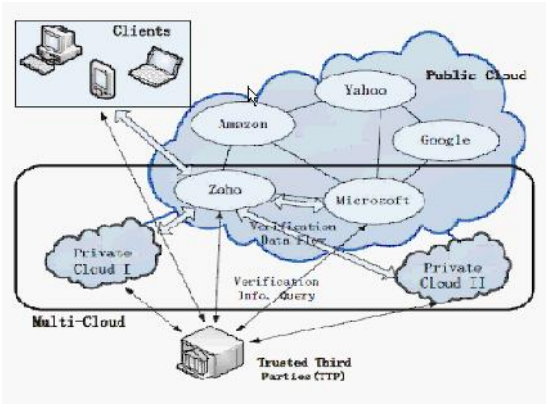


Fig-0, Cloud Architecture.

Provable data possession (PDP) [2] (or proofs of retrievability (POR) ) is such a probabilistic proof technique for a storage provider to prove the integrity and ownership of clients' data without downloading data. The proof-checking without downloading makes it especially important for large-size files and folders (typically including many clients' files) to check whether these data have been tampered with or deleted without downloading the latest version of data. Thus, it is able to replace traditional hash and signature functions in storage outsourcing. Various PDP schemes have been recently proposed, such as Scalable PDP and Dynamic PDP. However, these schemes mainly focus on PDP issues at untrusted servers in a *single* cloud storage provider and are not suitable for a multi-cloud environment.

## 2. EXISTING SYSTEM

There exist various tools and technologies for multicloud, such as Platform VM Orchestrator,

VMwarevSphere, and Ovirt. These tools help cloud providersconstruct a distributed cloud storage platform for managing clients' data. However, if such an importantplatform is vulnerable to security attacks, itwould bring irretrievable losses to the clients. For example, the confidential data in an enterprise may be illegally accessed through a remote interface provided by a multi-cloud, or relevant data and archives may be lost or tampered with when they are stored into an· uncertain storage pool outside the enterprise. Therefore, it is indispensable for cloud service providers to provide security techniques for managing their storage services.

## 3. PROPOSED SYSTEM

To check the availability and integrity of outsourced data in cloud storages, researchers have proposed two basic approaches called Provable Data Possession and Proofs of Retrievability .Ateniese et al. first proposed the PDP model for ensuring possession of files on untrusted storages and provided an RSA-based scheme for a static case that achieves the communication cost. They also proposed a publicly verifiable version, which allows anyone, not just the owner, to challenge the server for data possession..They proposed a lightweight PDP scheme based on cryptographic hash function and symmetric key encryption, but the servers can deceive the owners by using previous metadata or responses due to the lack of randomness in the challenges. The numbers of updates and challenges are limited and fixed in advance and users cannot perform block insertions anywhere.

## 4. DESIGN

### 4.1 Input Design

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

- What data should be given as input?
- How the data should be arranged or coded?
- The dialog to guide the operating personnel in providing input.
- Methods for preparing input validations and steps to follow when error occur.
-

### 4.3 Objectives

1.Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system.

2. It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities.
3.When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as when needed so that the user
will not be in maize of instant. Thus the objective of input design is to create an input layout that is easy to follow

### 4.2 Output Design

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making.
1. Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements.
2.Select methods for presenting information.

3.Create document, report, or other formats that contain information produced by the system.
The output form of an information system should accomplish one or more of the following objectives.

- Convey information about past activities, current status or projections of the
- Future.
- Signal important events, opportunities, problems, or warnings.
- Trigger an action.
- Confirm an action.

## 5. IMPLEMENTATION

### 5.1 Multi cloud storage

Distributed computing is used to refer to any large collaboration in which many individual personal computer owners allow some of their computer's processing time to be put at the service of a large problem. In our system the each cloud admin consist of data blocks . the cloud user upload the data into multi cloud. cloud computing environment is constructed based on open architectures and interfaces, it has the capability to incorporate multiple internal and/or external cloud services together to provide high interoperability. We call such a distributed cloud environment as a *multi-Cloud* .A multi-cloud allows clients to easily access his/her resources remotely through interfaces.

### 5.2 Cooperative PDP

**C**ooperative PDP (CPDP) schemes adopting zero-knowledge property and three-layered index hierarchy, respectively. In particular efficient method for selecting the optimal number of sectors in each block to minimize the computation costs of clients and storage service providers. cooperative PDP (CPDP) scheme without compromising data privacy based on modern cryptographic techniques.

### 5.3 Data Integrity

Data Integrity is very important in database operations in particular and Data warehousing and Business intelligence in general. Because Data Integrity ensured that data is of high quality, correct, consistent and accessible.

### 5.4 Third Party Auditor

Trusted Third Party (TTP) whois trusted to store verification parameters and offerpublic query services for these parameters. In our system the Trusted Third Party, view the user data blocks and uploaded to the distributed cloud. In distributed cloud environment each cloud has user data blocks. If any odification tried by cloud owner a alert is send to the Trused Third Party.

### 5.5 Cloud User

The Cloud Userwho have a large amount of data to be stored in multiple clouds and have the permissions to access and manipulate stored data.the User's Data is converted into data blocks . the data blocks is uploaded to the cloud. The TPA view the data blocks and Uploaded in multi cloud. The user can update the uploaded data. If the user wants to download their files, the data's in multi cloud is integrated and downloaded.

## 6. EXPERIMENTAL RESULTS

After the experiment some of the result has been mentioned as the screen shots for analyzing the data flow and security in data stored in the cloud can be observed carefully.


Fig-1, architectural outlook

Fig-2, user login to the cloud by registering

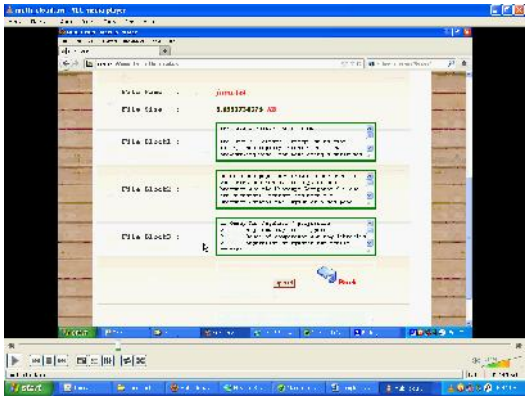

Fig-3, file stored by the user

Fig-4, file accessing by the user

Fig-5, monitoring system

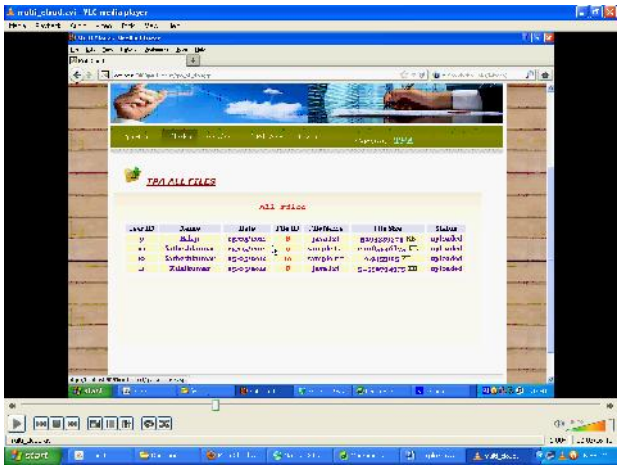

Fig-6, all files monitored during transfer.

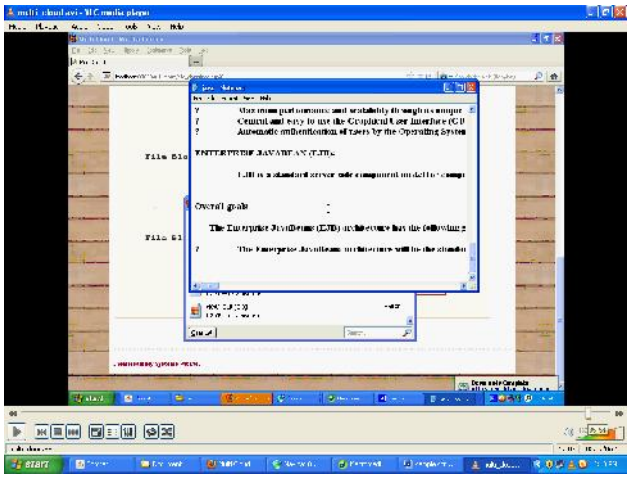

Fig-7, the software outlook

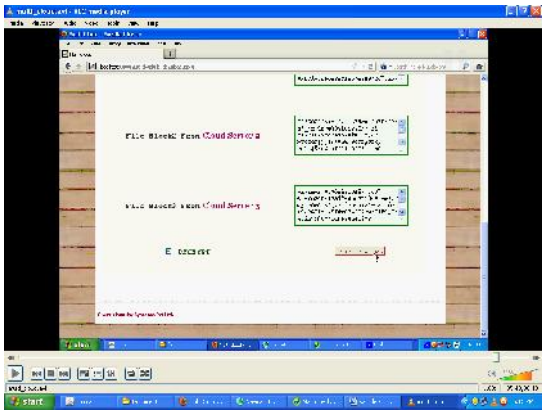

Fig-8, successful accessing and monitoring massage

**CONCLUSION**

We presented the construction of an efficient PDP scheme for distributed cloud storage. Based on homomorphic verifiable response and hash index hierarchy, we have proposed a cooperative PDP scheme to support dynamic scalability on multiple storage servers. We also showed that our scheme provided all security properties required by zeroknowledge interactive proof system, so that it can resist various attacks even if it is deployed as a public audit service in clouds. Furthermore, we optimized the probabilistic query and periodic verification to improve the audit performance. Our experiments clearly demonstrated that our approaches only introduce a small amount of computation and communication overheads. Therefore, our solution can be treated as a new candidate for data integrity verification in outsourcing data storage systems. As part of future work, we would extend our work to explore more effective CPDP constructions. Finally, it is still a challenging problem for the generation of tags with the length irrelevant to the size of data blocks. We would explore such a issue to provide the support of variable-length block verification.

**REFERENCES**

[1] B.Sotomayor,R.S.Montero,I.M.Llorente,andI.T.Foster, "Virtual infrastructure management in private and hybrid clouds," IEEE Internet Computing, vol. 13, no. 5, pp. 14–22, 2009.

[2] G. Ateniese, R. C. Burns, R. Curtmola, J. Herring, L. Kissner, Z. N. J. Peterson, and D. X. Song, "Provable data possession at untrusted stores," in ACM Conference on Computer and Communications Security,P.Ning,S.D.C.diVimercati,and P. F. Syverson, Eds. ACM, 2007, pp. 598–609.

[3] A. Juels and B. S. K. Jr., "Pors: proofs of retrievability for large files," in ACMConference on Computer and Communications Security, P. Ning, S. D. C. di Vimercati, and P. F. Syverson, Eds. ACM, 2007, pp. 584–597.

[4] G. Ateniese, R. D. Pietro, L. V. Mancini, and G. Tsudik, "Scalable and efficient provable data possession," in Proceedings of the 4th international conference on Security and privacy in communication netowrks, SecureComm, 2008, pp. 1–10.

[5] C.C.Erway,A.K¨upc¸¨u, C. Papamanthou, and R. Tamassia, "Dynamic provable data possession," in ACM Conference on Computer and Communications Security, E. Al-Shaer, S. Jha, and A. D. Keromytis, Eds. ACM, 2009, pp. 213–222.

[6] H. Shacham and B. Waters, "Compact proofs of retrievability," in ASIACRYPT, ser. Lecture Notes in Computer Science, J. Pieprzyk, Ed., vol. 5350. Springer, 2008, pp. 90–107.

[7] Q.Wang,C.Wang,J.Li,K.Ren,andW.Lou,"Enablingpublic verifiability and data dynamics for storage security in cloud computing," in ESORICS, ser. Lecture Notes in Computer Science, M. Backes and P. Ning, Eds., vol. 5789. Springer, 2009, pp. 355–370.

[8] Y. Zhu, H. Wang, Z. Hu, G.-J. Ahn, H. Hu, and S. S. Yau, "Dynamic audit services for integrity verification of outsourced storages in clouds," in SAC,W.C.Chu,W.E.Wong,M.J. Palakal, and C.-C. Hung, Eds. ACM, 2011, pp. 1550–1557.

[9] K. D. Bowers, A. Juels, and A. Oprea, "Hail: a high-availability and integrity layer for cloud storage," in ACM Conference on Computer and Communications Security, E. Al-Shaer, S. Jha, and A. D. Keromytis, Eds. ACM, 2009, pp. 187–198.

[10] Y. Dodis, S. P. Vadhan, and D. Wichs, "Proofs of retrievability via hardness amplification," in TCC,ser.LectureNotesin Computer Science, O. Reingold, Ed., vol. 5444. Springer, 2009, pp. 109–127.

[11] L. Fortnow, J. Rompel, and M. Sipser, "On the power of multiprover interactive protocols," in Theoretical Computer Science, 1988, pp. 156–161.

[12] Y. Zhu, H. Hu, G.-J. Ahn, Y. Han, and S. Chen, "Collaborative integrity verification in hybrid clouds," in IEEE Conference on the 7th International Conference on Collaborative Computing: Networking, Applications and Worksharing, CollaborateCom, Orlando, Florida, USA, October 15-18, 2011, pp. 197–206.

[13] Yan Zhu ; Beijing Key Lab. of Internet Security Technol., Peking Univ., Beijing, China ; Hongxin Hu ; Gail-Joon Ahn ; Mengyang Yu " Cooperative Provable Data Possession for Integrity Verification in Multicloud Storage "Parallel and Distributed Systems, IEEE Transactions on (Volume:23 , Issue: 12 )