Volume 3, Issue 3, March 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

# Data Integrity of Cloud Data Storages (CDSs) in Cloud

**Satyakshma Rawat**
Department of Computer Science Engineering
Amity School of Engineering and Technology
Noida-125

**Richa Chowdhary**
Department of Computer Science Engineering
Amity School of Engineering and Technology
Noida-125

**Dr. Abhay Bansal**
Head of the department
Computer Science and Engineering
Amity School of Engineering and Technology
Noida-125

***Abstract -- The growing need of Information Technology in every field has lead to the evolution of cloud computing for highly efficient usage of IT resources. Cloud provides services through internet. The computing power to the cloud environment is provided through collection of data centers or cloud data storages (CDSs) present at different location and connected by high speed networks. With the emergence of cloud computing the CDSs is also emerging. CDSs are an emerging technology and facing an important issue that is security. The integrity of data centers or CDSs is an extremely important issue. Our paper discusses the model based on MAS architecture of cloud and data encoding mechanism to enhance the integrity of Data centers or CDSs. Multi Agent Systems (MAS) are basically used in artificial intelligence area as a technique for finding solution to the problems. In cloud computing they are used to develop an architecture for integrity of the data present at data centers or CDSs. Data encoding is one of the basic mechanism of providing security, so we combine these two techniques to provide better integrity of data centers and data within the data centers.In this paper we will first provide an introduction about the cloud computing and methodologies which will be used by us. Then we will be giving the work we have done related to our proposed model and further describe about the future work in this area.***

***Keywords—Cloud Computing, data centers or data storages (CDSs), data integrity, CDSs integrity, MAS architecture, data encoding, CSPA, CDIBA***

## I. Introduction

Cloud computing is an emerging technology that is gaining fast acceptance day by day after Web2.0. Cloud computing is derived from Grid computing yet it's very different from it and provides service oriented applications. In this paper section 1.1 explains about what cloud computing is and what it deals with, section 1.2 deals with data centers and data storages (CDSs) and how they provide support for computing in cloud environment, section 1.3 discuss about data integrity and issues with integrity in cloud storages. The remaining part of the paper is organized as follows: section two describes the mechanisms currently imposed to ensure data integrity, section three gives Our proposed mechanism for data integrity, section four describes the future work related to data integrity and the paper is concluded finally in section five.

### 1.1 Cloud computing

Cloud Computing can be defined as a computing paradigm that provides dynamic computing environment for end users that is reliable and customized and also guarantees quality of service (QoS). Being synonymous to "Internet" it provides service oriented applications in form of infrastructure as a service (IaaS), platform as a service (PaaS), and software as a service (SaaS) over internet. Cloud computing is independent of Grid computing and Utility computing but when cloud computing is implemented publically then it is done as Utility computing. Cloud provides the facility of renting the services instead of buying or purchasing them and stores the complete data at its data centers for future access. The basic components of cloud computing are Clients, Services, Application, Platform, Storage and infrastructure. The essential characteristics a cloud must possess are:

- On-demand self service
- Broad Network Access
- Rapid Elasticity
- Resource Pooling
- Measured services

Cloud computing with its acceptance also has some growing needs which affect the complete working of cloud, and one of those needs is the need for "security". Cloud at present is lacking in its security needs in terms of data integrity, authorization and confidentiality. The clouds at present are being provided by specific vendors like Amazon and Google.

### 1.2 Data Centers or Cloud Data Storages( CDSs)

Data centers as the name suggests are the "house for data" for the purpose for data storage, management, analysis and dissemination. Data centers may exist in physical environment or virtually and can be organized as a public data center

*© 2013, IJARCSSE All Rights Reserved* 

for large scale usage or a private data center specific to an organization.

Data centers today are one of the main needs for the increasing information technology services and have an important role in cloud computing. The end-users provide their data to cloud to access it whenever required on the rental basis, therefore, the data provided is stored at data centers of cloud known as cloud data storages (CDSs). CDSs are present at different locations and store the complete data present on cloud. CDSs are also one of the rising trends in IT field and suffer from the issue of security within it. Even though there are many security issues related to Data centers or data storages but one of the most important issue is integrity of the data. The best examples of cloud storages (CDSs) are Amazon S3, Google Cloud Storage, iCloud by Apple, Google App Engine Blobstore, Windows Azure Storage, FilesAnywhere.

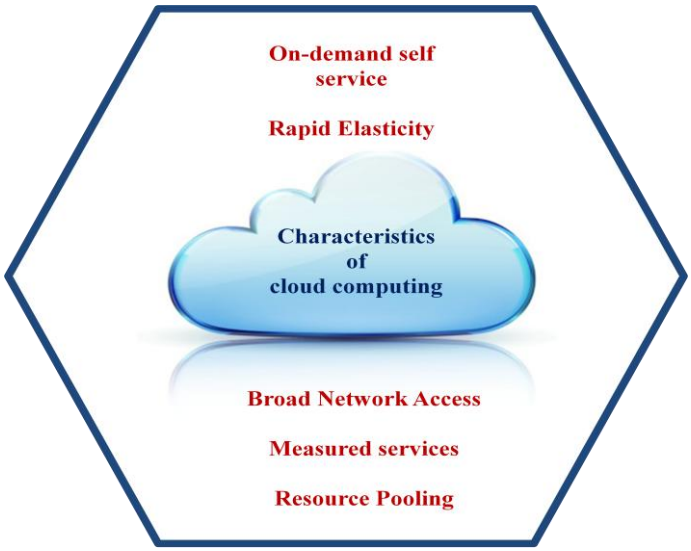


**Fig 1. The characteristics of cloud computing**

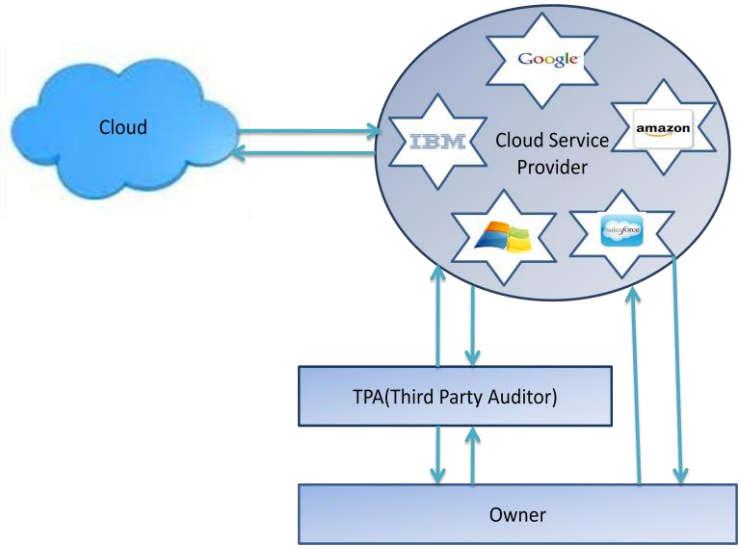


**Fig 2. Architecture of cloud data storage(CDSs)**

Cloud Storage(CDSs) doesn't have any specific or exact definition but it can be said as the mechanism of storing the end-user or client data in the remotely located cloud servers. Though Databases, Data warehouses and other traditional storage mechanisms provide high quality data storage but still cloud data storages are preferred over them as:

- The hosting providers or the companies don't need to install the storage devices physically in their datacenters.
- Storage management is the complete responsibility of the service provider and the hosting provider's company's focuses completely in their core business.
- Companies only need to pay for the storages they use and not the complete setup is to be established avoiding the complexities of data storages.

*© 2013, IJARCSSE All Rights Reserved* 

- The end-user doesn't need to keep backup but only store their data at service provider who looks after it completely.

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

**Manageability:** Cloud Data Storages (CDSs) basically focus on the maintenance of the large data which can't be stored on the end-user or clients system. This data stored in cloud storages are to be maintained and managed and for this reason cloud storages need to be self-managing to a large extent, so that reducing the maintenance trouble of data by the client.

**Availability:** Since the data being stored in cloud storage is very frequently accessed, therefore, there must be the proper availability of data always and no data corruption must occur as it is the back up of the complete information and data of client.

**Performance:** Since Cloud storages are used for cloud which are being used with internet because internet uses TCP for controlling the flow of data packets but TCP works for small size of data and is not suitable if the data size becomes larger, in such conditions cloud storages are used which provide large data storages and flow of data easily.

### 1.3 Data Integrity and Integrity of Cloud Data Storages(CDSs)

"Data Integrity" as the word in itself explains the 'completeness' or 'wholeness' of the data which is the basic requirement of the information technology. As Data Integrity is an essential in databases similarly integrity of Data Storages is an essential in the cloud, it is a major factor that affects on the performance of the cloud. The data integrity provides the validity of the data, assuring the consistency or regularity of the data. It is the complete mechanism of writing of the data in a reliable manner to the persistent data storages which can be retrieved in the same format without any changes later. As described above, in cloud, the complete storage of data provided by the end-user is done at the data centers or data storages (CDSs), and the security and integrity of the data lies on the vendor storing data in the data centers but not the cloud hosts. Cloud Storage is gaining popularity for the outsourcing of day-to-day management of data. Therefore integrity monitoring of data in cloud storages (CDS) is as essential for any data center, to avoid any data corruption or data crash. Data corruption or data failure can occur at any storage level. One of the most famous data failure occurred in Amazon in …. leading loss of complete client data stored in it. Therefore just storing data at cloud data storages or data centers doesn't ensure the integrity of data, but some mechanisms are to be implemented at each storage level to ensure the data integrity. Data Integrity is most important of all the security issues in cloud data storages because it not only ensures completeness of data but also ensures that the data is correct, accessible, consistent and of high quality.

## II. Literature Review

In Cloud Computing there are many threats which are avoiding the wide acceptance of cloud as explained above. One of the major threats is data privacy and data integrity in cloud storages. There is lot of research going on in this field to ensure and provide data integrity in cloud storages. 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 implementation 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] proposed a scheme to ensure data reliability and retrievability of data for multiple servers. Many mechanisms has been proposed till now to guarantee and ensure complete data integrity and data privacy of cloud storages based on encryption and cryptographic mechanisms using hash values and data encoding. Filho[9] proposed a RSA-based hash data integrity mechanism for peer-to-peer file sharing networks and exponentiation of complete data file is done, but this mechanism can be followed for the files and data of large size, and also this mechanism focuses on the static data files and not on files being dynamic in nature having localization problem. Atan and Abdullah[10] proposed a data integrity mechanism of Cloud Zone which uses the concept of Multi Agent System(MAS) architecture. MAS are the techniques used in artificial intelligence where they communicate with each other to find a complete solution. All the mechanisms proposed till now and discussed above "provide the services in cloud storages (CDSs) for ensuring the integrity of data transmission and not the integrity problem of CDS". MAS based system provides integrity of CDS. Since it provides the integrity of CDS, this mechanism provide reliable retrievable of data form CDS, but not the reliable data transmission.

## III. Proposed Methodology

As discussed above the integrity within cloud storage can be of two things, that is, integrity of data being transmitted from CDS and integrity of CDS. The mechanism for providing both type of integrity are present, and mechanism for providing integration together are also present but here we are discussing the most two commonly used methodology for

*© 2013, IJARCSSE All Rights Reserved* 

integration assurance and then will introduce our model combining these two mechanisms to provide both kind of integrity together at the same time.

**Multi Agent System (MAS) Architecture for CDS Integrity:**

MAS architecture is a mechanism being developed from the concept of Multi Agent system(MAS) in artificial intelligence defined as "loosely-coupled network of entities that work together to find solutions for the problems which are beyond the knowledge of single entity". It is implemented on basic cloud architecture and consists of two main layers as cloud resource layer (cloud server side) and MAS architecture layer(cloud client-side). At cloud resource layer as the name suggest consists of all the resources of cloud like storage servers and application servers which provide a platform or power to CDS. MAS Architecture layer known as cloud zone consists of 5 agents but most widely used agents for integrity are Cloud Service Provider Agent(CSPA) and Cloud Data Integrity Backup Agent(CDIBA).

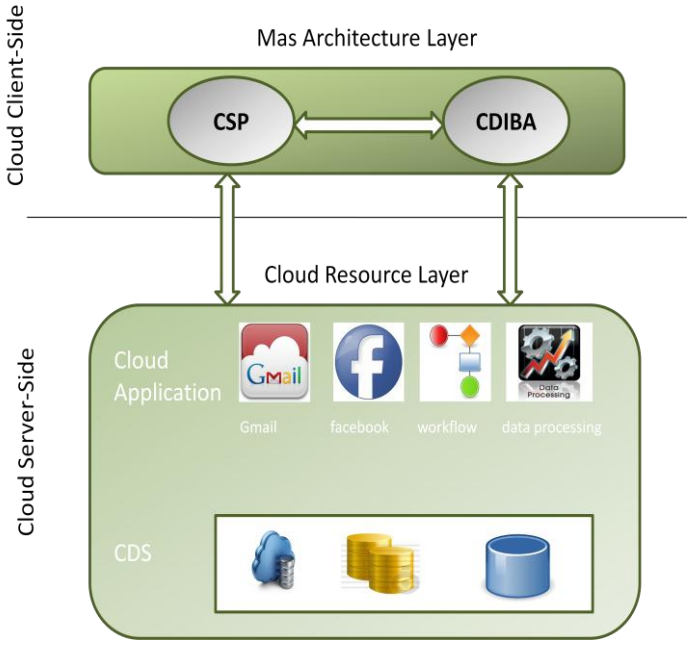


**Fig 3. MAS Architecture "CLOUD ZONE"**

To provide integrity of CDS the MAS architecture implies some of the security policies using CSPA agent whose main responsibility is to backup data from the Cloud Zone and send regular security alerts or alarms whenever there is a human error when data for cloud is entered, a software bugs or viruses occur, some hardware fault like system crash, or error during the transmission of cloud data from one computer to another. The MAS architecture uses Prometheus methodology for designing the prototype and Java Agent Development Framework Security (JADE-S) for its implementation.

**Data Encoding Mechanism**

Data encoding is one of the basic methodologies used for the purpose of integration of data being transmitted. Being one of the first security mechanisms provided for network security, it is still a most widely used mechanism with regular changes and addition in its working. The most basic mechanism of data encoding used in CDS for data integration is based on hash values, that is, encrypting data using encryption mechanism and then using hash values at server and client side to check the integrity of data being transmitted. This mechanism provide assurance of reliable data transfer but not reliable retrievable of data. Many mechanisms as discussed above have been proposed till date to ensure integrity in best and effective manner and lacks at some point leaving a space for further studies on this issue.

**Our Mechanism**

As discussed above the two mechanisms for integrity of CDS and integrity of data transmission. We in our methodology are proposing a mechanism where these two mechanisms are combined, that is, MAS architecture for CDS and data encoding using hash values are combined together to give a new mechanism.

MAS architecture basically uses two agents in client side layer for data integrity, CSPA which takes care of all the backing up of data in cloud zone and generate regular alarms if any error occurs, but doesn't guarantee any authorization as it only talks about the data which is being entered into cloud storage ad not about the data accessed by the clients regularly, that is, if there is any unauthorized access to data and data is being entered in correct format, then the cloud will provide the data.

So to ensure that the data being entered is integrated and data being accessed from cloud is also integrated, we are combining these two mechanisms. This can be done inserting a hash value concept in CDIBA agent of MAS architecture. CDIBA is responsible for the maintenance of cloud storage when data is entered into it, and if the data goes out of the

*© 2013, IJARCSSE All Rights Reserved* 

cloud storage the hash values being used can be used to verify the data being transmitted is in correct format. Hence the complete process of reliable retrieval and reliable data transmission is guaranteed at the same time.

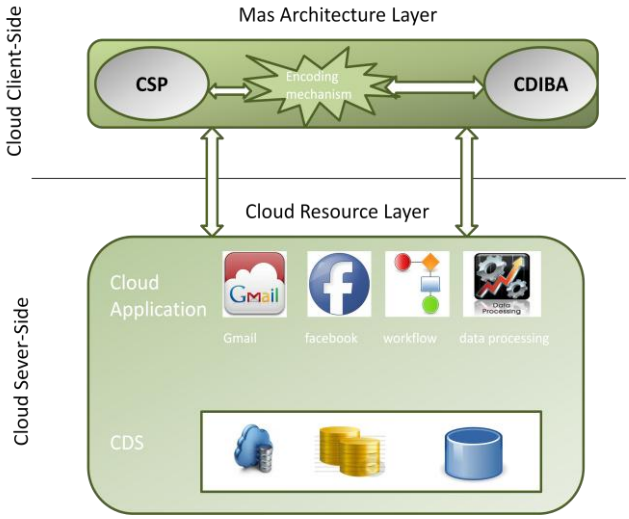


**Fig 4. Encoded MAS Architecture**

## IV. Conclusion And Future Work

In this paper we discussed about cloud computing and the role of cloud storages (CDSs) in cloud computing, and describing the most important security threat of CDS which is data integrity and data privacy, the proposed mechanisms for integrity assurance and the problems being faced in these mechanism. Later we discussed about MAS architecture and data encoding using hash values for the purpose of integrating data transmission as well data integrity of CDS by encrypting and hashing the data in CDS using CDIBA agent of Cloud Zone.

Security is an issue which always has some work to do because with rising security, the breaking points in security also occur. The mechanism we proposed can be implemented using much better encoding mechanism so that the security rises more and data integrity enhances more and more.

**Acknowledgement**

We would like to thank our mentor Dr. Abhay Bansal for motivating us to pursue the research work in this area. He has always been there to educate us, guide us, motivate us and enlighten us. This would have not been possible without him.

**Refrences**

[1] 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.

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

[3] K. D. Bowers, A. Juels, and A. Oprea, "Proofs of Retrievability: Theory and Implementation," Cryptology ePrint Archive, Report 2008/175,2008

[4] K. D. Bowers, A. Juels, and A. Oprea, "HAIL: A High-Availability and Integrity Layer for Cloud Storage," Cryptology ePrint Archive, Report 2008/489, 2008.

[5] Jia Xu and Ee-Chien Chang, "Towards efficient proofs of retrievibility in cloud storage".

[6] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson,and D. Song, "Provable Data Possession at Untrusted Stores,"In Proceedings.Of CCS '07, pp. 598–609, 2007.

[7] G. Ateniese, R. D. Pietro, L. V. Mancini, and G. Tsudik, "Scalable and Efficient Provable Data Possession," In Proceedings of SecureComm '08, pp. 1–10, 2008.

[8] R. Curtmola, O. Khan, R. Burns, and G. Ateniese, "MR-PDP: Multiple-Replica Provable Data Possession," In Proceedings of ICDCS '08, pp.411–420,2008.

[9] D.L. Gazzoni Filho, and P. Barreto, "Demonstrating Data Possession and Uncheatable Data Transfer," Book Demonstrating data possession and uncheatable data transfer, Series Demonstrating data possession and uncheatable data transfer, ed., Editor ed.^eds., Citeseer, 2006, pp. 150.

[10] A.M. Talib, R.Atan, R.Abdullah and M.A. Azmi Murad, "CloudZone: Towards an integrity layer of cloud data storage based on Multi Agent System Architecture", 2011 IEEE Conference on open systems(ICOS 2011), September 25-28 2011.

*© 2013, IJARCSSE All Rights Reserved* 