Enabling Data Integrity Check and Corruption Prevention in Thin Cloud Storage

N. Kamaladheepan, A. Marimuthu

Government Arts College, Coimbatore

Summary
Cloud Computing has evolved and matured, it gets growing interest in the enterprise market where economic pressures are challenging traditional IT operations. Many IT organizations face inefficiency in areas like funding projects, resource utilization, manual provisioning times, and organizational silos. Cloud Computing is focused on addressing these issues by cutting costs through better standardization, higher utilization, greater agility, and faster responsiveness of IT services . A main concern on Cloud journey is security of the infrastructure and the information stored in the infrastructure. To support these requirements, most organizations emphasis to move from maintaining tied infrastructure to a loosely coupled service oriented model. Data integrity became one critical factor. This paper aims to give solution to protect cloud data by dividing and storing the data in encrypted form. The data integrity is checked by the client and corrupted data is regenerated. This enables high data integrity in cloud storage.

Key words
Cloud computing, Data Integrity, Multi-server, Data Corruption, Regeneration

1. Introduction

Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable and reliable computing resources (e.g., networks, servers, and minimal consumer management effort or service provider interaction.
Cloud computing isn't much a single technology as it is a combination of many storage, applications, services) that can be rapidly provisioned and released with existing technologies. The essential characteristics of cloud computing are on-demand self-service, ubiquitous network access, resource pooling, location independence, rapid elasticity and measured service.

2. Cloud Computing Types

There are several service and deployment models for implementing cloud technology. Each has
its advantages and disadvantages with significant implications for any organization researching or actively considering a cloud deployment. The service models are

Cloud Software as a Service (SaaS) - Use provider's applications over a network
Cloud Platform as a Service (PaaS) - Deploy customer-created applications to a cloud.
Cloud Infrastructure as a Service (IaaS) - Rent processing, storage, network capacity, and other fundamental computing resources.[3]
The deployment models, which can be either internally or externally implemented, are
Public cloud- Sold to the public, mega-sale infrastructure
Private cloud- Enterprise owned or leased
Hybrid cloud- Composition of two or more clouds
Community cloud- Shared infrastructure for specific community [4]

3. Challenges In Cloud Data Storage

Cloud storage offers an on-demand data outsourcing service model, and is gaining popularity due to its elasticity and low maintenance cost. However, security concerns arise when data storage is outsourced to third-party cloud storage providers. Integrity is the guarantee that the message sent is the message received and that the message is not intentionally or unintentionally altered.
It is desirable to enable cloud clients to verify the integrity of their outsourced data, in case their data have been accidentally corrupted or maliciously compromised by insider/outsider attacks. Some of the elements used to ensure integrity are firewall services, communication security management and intrusion detection services. [1]

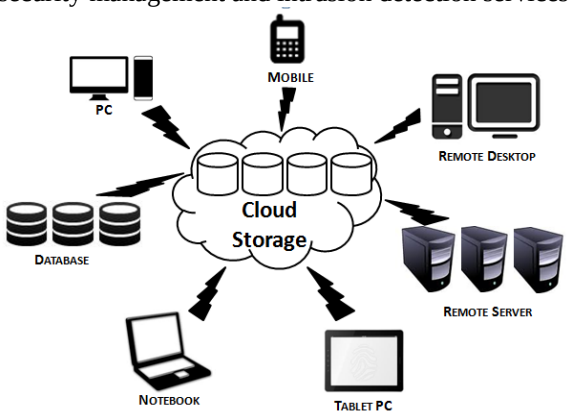

Fig.1 Cloud Storage System

Manuscript received March 5, 2016
Manuscript revised March 20, 2016

4.Approaches To Secure Cloud Data

This paper mainly deals with protecting the long term archival data in the thin cloud storage. Thin cloud storage is that the client only uses the services of the vendor by paying to the vendor. Here, the client uses the cloud storage for achieving the data in the server. As the data may not be accessed frequently by the client, there may be more chance of the attackers to corrupt the data in the cloud storage. A new Data Integrity Protection is proposed to protect the data in cloud.

Also the client can check the integrity of data by randomly picking up the data and checking it using Trusted Party Auditor (TPA). TPA is an agent which is responsible for checking the originality of the client data. If the client data is corrupted, then TPA notifies the same to the client.

A regeneration erasure code mechanism is proposed to regenerate the corrupted data from the replica servers. In prior methods, solution for this problem are proposed for single server only. In single server case, if the server fails then the whole data gets lost.

Suppose that we outsource storage to a server, which could be a storage site or a cloud-storage provider. If we detect corruptions in our outsourced data (e.g., when a server crashes or is compromised), then we should repair the corrupted data and restore the original data. However, putting all data in a single server is susceptible to the single point- of-failure problem and vendor lock-ins. A plausible solution is to stripe data across multiple servers. In this research work, the solution is proposed for multi-server setting in which if, data corrupted by attackers from a cloud server can be reconstructed from the secondary servers. Thus, to repair a failed server, we can

(i) Read data from the other surviving servers,

(ii) Reconstruct the corrupted data of the failed server, and

(iii) Write the reconstructed data to a new server.

In particular, erasure coding has a lower storage overhead than replication under the same fault tolerance level. In a distributed environment, an attacker chooses a specific client but the distribution of data into multiple server makes the attacker's job more difficult. Data is encrypted and divided in to chunks. If a part of data is corrupted in a server, then it is recovered from the secondary server.

MR-PDP and HAIL [2] extend integrity checks to a multi-server setting using replication and erasure coding, respectively.

5. Data Integrity In Thin Cloud Storage

The basic operations to be performed for storing the file in multiple servers are

(i) Generate the secret key that are used for encrypting and decrypting the files.

(ii) Encode the file F of size |F| into n pieces of size |F|/k each, where k<n

(iii) Split the file in to chunks and apply AECC for the i^{th} chunk P*i*.

(iv) Apply XOR operation for P_1i *XOR* P_2i, P_2i *XOR* P_3i,…… P_ni

(v) Upload the code chunks P*i*'s to respective servers

To maintain the integrity of data in servers, we verify random chunks chosen from the servers. Thus k(n-k) code chunks from any k servers can be decoded to the original k(n-k) native chunks, we must have rank (A)=k(n-k). Now we can pick a chunk Pi from one of the remaining n-k servers and its rank is also checked. If it is not consistent, then we have to reconstruct the data. [1]

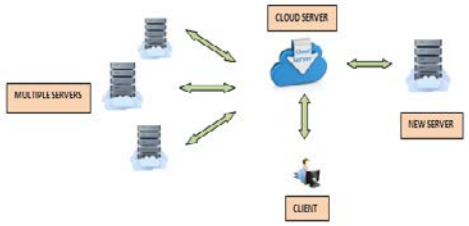

Fig.2 Multi-server Environment

Download the corrupted file, its AECC parities from the server and apply error correction. Verify the correct chunk with MAC. Repeat the same for (n-k) code chunks.

If the server fails or having a large number of corrupted data, then download the chunk from all remaining n-1 servers. Again encode and upload the data to the new server.

Based on the algorithm, the below steps are proposed to check the data integrity in cloud storage.

(i) The client who wants to store data in the cloud should register the details such as user name, password, email id, phone number.

(ii) Once the registration is completed, the client is allowed to upload file to the cloud server.

(iii) The NCC cloud connector is used to connect to the Dropbox public cloud and its space is utilized.

(iv) The file being uploaded is encrypted using Advanced Encryption Standard (AES-128 bit algorithm) and divided in to chunks, parity bit is added and stored in multiple cloud servers.

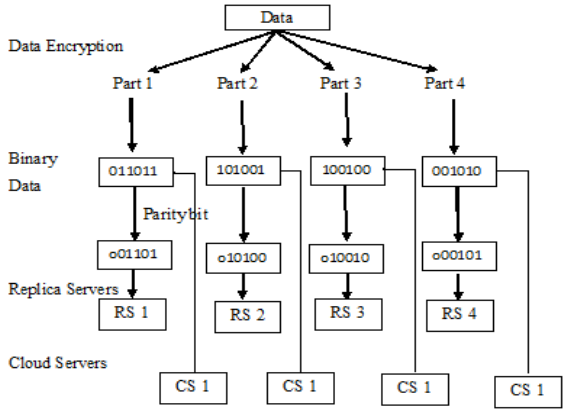

Fig.3 Parity Bit Addition and Erasure Code

(v) Trusted Parity Auditor (TPA) uses the parity bits in the primary cloud servers for integrity check.
(vi) Encrypted data without parity addition is stored in the replica servers.
(vii) Client performs the data integrity check on randomly chosen parts of data in the cloud server using TPA.
(viii) The TPA uses SHA 256 algorithm and compares the hash value of the corrupted file and the original file in the replica server. If it is not equal, then an intimation is sent to the client mail about the data corruption.

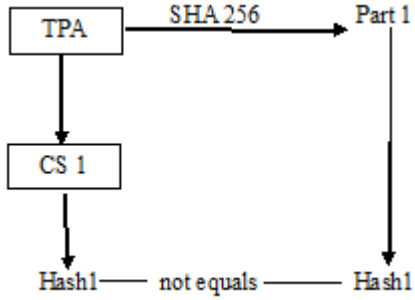

Fig.4 Data Integrity Check using SHA Algorithm

(ix) While downloading the data, all the data are retrieved from the replica server without data loss.
(x) As a protective mechanism, all the parts of data are combined with XOR operation and stored in backup server in a different location. This enables high data integrity in cloud.

6. Analysis And Evaluation

The running time of Data Integrity codes were evaluated in the local cloud platform. We measure the running time or each operation. Our results are averaged over 10 runs. We used files of size 100MB for evaluation purposes.

Using Secure Hash Algorithm, the integrity of the data are verified. Time is consumed according to the size of the data encoded. Time taken for all data are displayed.

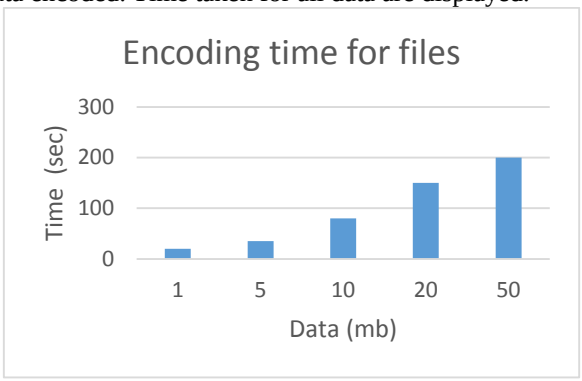

Fig.5 Time taken by files for encoding

Also the ranks of different chunks are checked and evaluated. The below graph shows the time taken to check different percentage size of files.

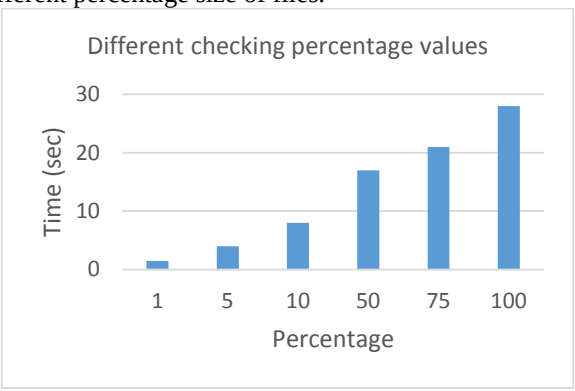

Fig.6 Different checking percentage values

7. Conclusion

In this paper, we have outlined the new approach to perform secure replication of stored information. This approach enables the client to verify the integrity of their data in cloud. This is a dominant technique which will provide better results for security and availability of data. We can use this secure replication technique in order to build a secure and reliable distributed storage. We expect the enhancement done in this technique will increase the quality by different data mart host with cloud provider and store information accordingly based on sensitivity. This new approach can be used by different cloud providers and other organizations.

References

[1] Henry C. H. and Patrick P. C. Lee, "Enabling Data Integrity Protection in Regenerating-Coding-Based Cloud Storage", 31st International Symposium on Reliable Distributed Systems, IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 25, NO. 2, FEBRUARY 2014

[2] K. Bowers, A. Juels, and A. Oprea, "HAIL: A High-Availability and Integrity Layer for Cloud Storage," Proc. 16th ACM Conf.Computer and Comm. Security (CCS '09), 2009.H. Abu-Libdeh, L. Princehouse, and H. Weatherspoon, "RACS: A Case for Cloud Storage Diversity," Proc. First ACM Symp. Cloud Computing (SoCC '10), 2010.

[3] M. Armbrust, A. Fox, R. Griffith, A.D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "A View of Cloud Computing," Comm. ACM, vol. 53, no. 4, pp 50-58, 2010.

[4] Cloud Security- A comprehensive guide to secure cloud computing by Ronald L. Krutz, Russell Dean Vines, Wiley India Pvt. Ltd. Edition 2010, ISBN: 978-81-265-2809-7

[5] Cloud Computing by David cookers, Tata McGraw Hill 2012 edition, ISBN-13:978-1-25-906104-2

[6] Gupta Sarika, Sangita Rani Satapathy, Mehta Piyush and Tripathy Anupam, "A Secure and Searchable Data Storage in Cloud Computing", 3rd IEEE International Advance Computing Conference (IACC), 2013, page 106-109.

[7] Taeho Jung, Xiang-Yang, Zhiguo Wan, Meng Wan, "Privacy Preserving Cloud Data Access With Multi-Authorities", Proceedings IEEE INFOCOM, 2013, page 2625-2633.

[8] Amazon.com. Amazon simple storage service (Amazon S3), 2008. Referenced 2008 at aws.amazon.com/s3.