International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 07 | July 2020 www.irjet.net p-ISSN: 2395-0072

STORESIM: INFORMATION LEAKAGE AWARE MULTICLOUD STORAGE SYSTEM

NANBON ABERA TOLASA[1]

[1]*Master of Technology, Department of Software Engineering, Jawaharlal Nehru Technological University Hyderabad (JNTUH), School of Information Technology, Hyderabad, Telangana*

Abstract - *Many schemes have been recently advanced for storing data on multiple clouds. Distributing data over different cloud storage providers (CSPs) automatically provides users with a certain degree of information leakage control, for no single point of attack can leak all the information. However, unplanned distribution of data chunks can lead to high information disclosure even while using multiple clouds. In this paper, I study an important information leakage problem caused by unplanned data distribution in multicloud storage services. Then, I presented* ***StoreSim****, an information leakage aware storage system in multicloud. StoreSim aims to store syntactically similar data on the same cloud, thus minimizing the user's information leakage across multiple clouds. I designed an approximate algorithm to efficiently generate similarity-preserving signatures for data chunks based on MinHash and also design a function to compute the information leakage based on these signatures.*

Lastly, provided optimal information leakage storage system based on planned distribution of data chunks across multiple clouds.

1. INTRODUCTION

1.1 Introduction

Now-a-days technology is growing and everyone is using different devices such as mobile phones, tablets and computers to store their massive critical data. There are many cloud storage providers such as Microsoft OneDrive, iCloud, and Dropbox which used to store users data over the cloud. These storage services have good demands due to the simplicity and cheap storage price. However, these storage providers are taking the control of user's data which may leak the user's data by different reasons such as trap door, hack bribe and coercion. [1]

The proper way to reduce the degree of information loss is using multiple clouds, which used to reduce one point failure in single cloud. The recent cloud storage providers like, Dropbox is used rsync-like protocols to operate the local file to remote file in their storage. In rsync-like protocols each user file is divided into chunks and hashed with fingerprinting algorithms such as SHA-1, MD5. Hence when a local file is modified, the changed hash will be uploading to the cloud. In fact, now a days service providers like Dropbox, Google Drive are used data deduplication methods to check similarity between data chunks by their fingerprints, but this fingerprint will check only as data nodes are duplicate or not. It is simple to check identical chunks, but to efficiently find out similarity between chunks is a complex task due to lack of similarity preserving signatures. [6]

Therefore, to address this problem I provided StoreSim which aware information leakage storage system by storing similar data to same cloud and I designed MinHash algorithm to efficiently generate similarity-preventing signatures for data chunks and designed function to control information leakage.

1.2 Objective of the project

Now-a-days multicloud storage services are very popular and many of cloud storage providers such as Google Drive, Drobox, and Amazon S3 are given a services for users to store the data's. [6] However, unplanned distributions of data on multicloud storage providers would produce high degree of information leakage and one point failure in multiple clouds. Hence, the objective of this project is to find out the optimal techniques to control information leakage in multicloud. Then, I presented StoreSim which is information leakage aware storage system and store the similar data to same cloud based on jaccard similarity. Therefore, I provided optimal information leakage in multicloud storage system which used to store users data in efficient and secure manner.

3. OVERVIEW OF THE SYSTEM

3.1 Existing System

In our daily tasks, we are storing our critical data over different cloud storage providers such as Google Drive, Dropbox, and iCloud, but unplanned distribution of data chunks over multiple clouds storage providers will produce one point failure in cloud data. [6]

In the existing system, data deduplication techniques are adopted during distributing data over multiple clouds, which used to identifies the same data chunks by their

fingerprints which generated by fingerprinting algorithms such as SHA-1, MD5, but any change to the data will produce a different fingerprint. However, these fingerprints can only detect whether or not the data nodes are duplicate, which is only good for exact equality testing. [1] [6] Determining the similarity of data chunks is relatively simple, but it is difficult tasks to determine efficiently due to the absence of similarity preserving signatures.

3.1.1 Disadvantages of Existing System

- ✓ Redundancy of similar data during frequently modification of multiple clouds data.
- ✓ No enough security.
- ✓ There is a probability of all information to be loss.

3.2 Proposed System

To minimize the information leakage during unplanned distributions of data, I presented StoreSim system, which aware an information leakage storage system in the multicloud. It used to store similar data over the same cloud syntactically to minimize losing of information in multiple clouds. This system is achieved their goal by using novel MinHash algorithm, to efficiently generate similarity-preventing signatures for data chunks and designed a function to compute the information leakage based on these signatures.which given by fingerprint algorithms such as SHA-1,MD5. Now a user's can store data in multiple clouds, modify each cloud data as they need and download the data from multiple clouds in secure manner.

Therefore, I developed advanced system which used to store data in secure, effective and efficient manner over centralized multicloud storage system to minimize information leakage.

3.2.1 Advantages of Proposed System

- ✓ Each data chunk which stored across multiple clouds is encrypted and no one can see without decryption.
- ✓ There is no vendor-lock-in problem and user can stored his multiple cloud data easily for along a period of time continuously.
- ✓ There are cost optimization, data consistency and availability in the proposed system;
- ✓ Reduce the chance to loss all the information in the same time.

3.3 System Modules

In this project work, I used three modules and each module has own functions, such as:

1. Client module
2. Metadata Servers module
3. Cloud Service Providers module

3.3.1 Client module

This is in charge of pre-processing the users' data for the purpose of optimization, such as chunking (i.e., dividing files into individual chunks of a maximum size data unit), deduplication (i.e., avoiding storing and re-transmitting the same content already available on the remote servers), delta encoding (i.e., transmission of only modified portions of a file), bundling (i.e., the transmission of multiple small files as a single object) and encryption/decryption;

3.3.2 Metadata servers module

These are used to store the metadata database about the information of files, CSPs and users, which usually are structured data representing the whole cloud file system;

3.3.3 Cloud Service Providers module

In this module, I presented the cloud functionality. The cloud user must login to the page to views files, view requests and approve the requests which sending from other users or owner of data to get both cloud keys. I used DriveHQ cloud service which integrated to the local system to store different data chunks after chunking.

4. SYSTEM ARCHITECTURE

In this project work, I used three tier architecture which represent the flow of requests between users, database and servers. In this architecture the layers such as presentation layer, business layer, and data link layer are separated.

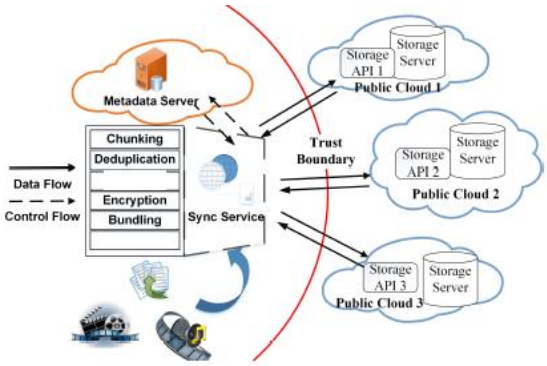

Fig- 4. 1 System Architecture

Advantage of three tier architecture

1. High degree of flexibility in deployment platform and configuration.

2. Better re-use and high performance

3. Improve data integrity and security Scalability and etc.

5. RESULTS

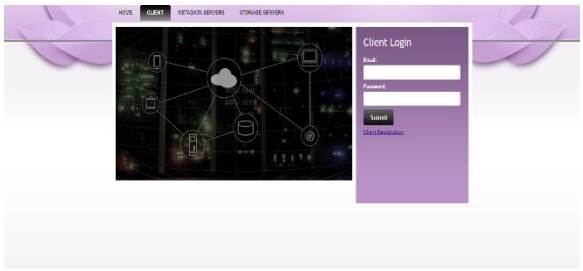

Fig- 5.1: Client Login

Fig -5.2: File Upload

Fig -5.3: File Split

Fig -5.4: Request to Download

Fig -5.5: Approve Request

6. CONCLUSION

Most of the people are storing their data over the multiple clouds for security purpose. However, unplanned distribution of data over the multiple clouds may produce the probability to loss all data, which increase degree of information leakage in multicloud. To optimize information leakage, I provided StoreSim system, which is a storage system, that aware information loss in multiple clouds and storing the similar data to same cloud. I designed MinHash algorithm to efficiently generate similarity-preserving signatures for data chunks and designed a function to compute the information leakage based on these signatures which are hashed by fingerprint algorithms such as SHA-1, MD5.

Finally, I provided the optimal multicloud storage providers which used to minimize information leakage efficiently.

Future Enhancement

- ✓ Adding user password update option.
- ✓ As a security technology updated, then system security will also update.
- ✓ Adding the number of clouds more than two to control information leakage in advanced way.

7. REFERENCES

[1] H. Chen, Y. Hu, P. Lee, and Y. Tang, "Nccloud: A network-coding-based storage system in a cloud-of-clouds," 2013.

[2] T. G. Papaioannou, N. Bonvin, and K. Aberer, "Scalia: an adaptive scheme for efficient multi-cloud storage," in Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis. IEEE Computer Society Press, 2012, p. 20.

[3] Z. Wu, M. Butkiewicz, D. Perkins, E. Katz-Bassett, and H. V. Madhyastha, "Spanstore: Cost-effective geo-replicated storage spanning multiple cloud services," in Proceedings

of the Twenty-Fourth ACM Symposium on Operating Systems Principles. ACM, 2013, pp. 292–308.

[4] U. Manber et al., "Finding similar files in a large file system." in UsenixWinter, vol. 94, 1994, pp. 1–10.

[5] P. Mahajan, S. Setty, S. Lee, A. Clement, L. Alvisi, M. Dahlin, and M.Walfish,"Depot: Cloud storage with minimal trust," ACM Transactions on Computer Systems (TOCS), vol. 29, no. 4, p. 12, 2011.

[6] J.-M. Bohli, N. Gruschka, M. Jensen, L. L. Iacono, and N. Marnau, "Security and privacy-enhancing multicloud architectures," Dependable and Secure Computing, IEEE Transactions on, vol. 10, no. 4, pp. 212–224, 2013.