Dynamic and Secure Deduplication towards Encrypted Data in Big Data Cloud Service

Dr. K.Kavitha

Assistant Professor, Department of Computer Science
Mother Teresa Women's University, Kodaikanal

***Abstract* - Data Storage Services are the most popular and significant task in Cloud Service. Here data's are stored on Cloud in an encrypted form. Those data's also meet some challenges for data deduplication because of storage inefficiency and secure privacy. Many researchers making observations have offered different techniques for supporting duplication effectively, but still it cannot support the facts way in control readily able to make adjustment ie) data access control flexibly. In this paper, the author reviewed two important concepts such as 2FA-OTP and dynamic deduplication for improvising ownership challenges and privacy and also proposed an integrated technique for performing deduplication with more secure. The proposed integrated approach provides high security in privacy policy and improves storage efficiency using deduplication techniques.**

Keyword: Deduplication, encryption, Proof of ownership, Cloud Storage, Quality of Service

I. INTRODUCTION

Cloud computing is an emerging technology for utility computing system. The concept of cloud provides computing resources as a utility or a service on demand to customers over the Internet [1]. Cloud providers pool computing resources together to serve customers via a multi-tenant model. Computing resources are delivered over the Internet where customers can access through various client platforms [3]. Customers can access the resources on-demand at any time with the cloud provider. From a customers' point of view, computing resources are infinite, and customer demands can rapidly change to meet business objectives. This is facilitated by the ability for cloud services to scale resources up and down on demand leveraging the power of virtualization.

Cloud storage is the main services in cloud computing which provides virtualized storage on demand to customers. As the amount of data in the cloud increases, customers expect to reach the on-demand cloud services at any time, while providers are required to maintain system availability. Providers need to dramatically reduce data volumes, so they can reduce costs while saving energy consumption for running large storage systems. Data deduplication is a technique whose objective is to improve storage efficiency. Deduplication reduces both storage space and network bandwidth[7]. Many approaches and techniques have been proposed to achieve storage efficiency and also improve its fault tolerance. These techniques provide redundancy of data chunks after performing deduplication. Current data deduplication mechanisms in cloud storage are static schemes. Data usage in cloud changes overtime; some data chunks may be read frequently in a period of time, but may not be used in another period. Due to the drawback of static schemes, deduplication in cloud storages requires a dynamic scheme which has the ability to adapt to various access patterns and changing user behavior in cloud storages.

The contribution of this paper is a dynamic data deduplication scheme and OTP for efficient cloud storage and secure privacy, in order to fulfill a balance between storage efficiency and also to improve performance in cloud storage systems. The rest of this paper is organized as follows: section II presents background concepts and related work. Section III demonstrates a proposed system model. Section IV concludes this paper.

II. RELATED WORK

Dynamic Data Duplication in Cloud Storage:

The existing system[11] focuses Client side deduplication using file hashing technique. Hashing process is performed and connects to any deduplication according to their loads. Deduplicator identifies the duplication based on hash value in Meta data server. Some systems may keep a number of copies of each file with static number. However files with many references require more duplication. To solve this issue and improve availability, author proposed dynamic data duplication technique considers dynamicity and QoS (Quality of Service) of the Cloud environment. In this system model, after identifying duplication, redundancy manager calculates optimal number of copies for the corresponding file based on number of references and level of QoS necessary. Here numbers of copies are dynamically changed based on the changing number of references, QoS level and demand of file which automatically improves storage efficiency.

DOI: 10.21817/ijet/2017/v9i5/170905317

Secure and Efficient cloud storage Data Deduplication System:

Security and Privacy are the major concern for Public Cloud. To address the security challenges, author[12] made an attempt to formalize the notion of secured and efficient Cloud storage system. Client side deduplication scheme for security, storing and sharing developed for hybrid cloud. Two factor Authentication scheme of user along with POS of files to address the problem of authorized data deduplication, in which tokens are generated by private cloud server with private keys. This method provides double security through OTP technique.

III. PROPOSED METHODOLOGY

This paper proposed the further extension of existing system by integrating security and dynamicity. Here, token generated at private cloud by including PoW with 2FA-OTP[12] for enhancing security and maintain dynamicity and QoS[11] for improving availability while maintaining storage efficiently.The proposed scheme contains the following main aspects

1. User Authentication

User enters authenticated ID and PWD for availing Cloud services. Ownership Verification Protocol has been applied to ensure that it is the original authenticated person that avails their Cloud space for possessing/uploading data.

2. Data Upload

If user tries to upload data, token request will be sent to private cloud. Private Cloud generates the token and sends through OTP to ensure the Security and Privacy of data

3. Data Deduplication

If User tries to store the same data which has already stored in Cloud Service Provider, the data duplication occurs. In order to minimize the storage space effectively, deduplication checking process will be performed.

if deduplication check is Negative -- **Encrypted Data**

If deduplication check is Negative, the data holder encrypts the data using randomly selected symmetric key to ensure the security and stores the data at Cloud Service Provider along with generated token**.**

if deduplication check is Positive – **Redundancy Checking**

If deduplication occurs then check the number of redundancy occurs based on QoS Manager Procedure which creates a file pointer for reference and stored it in CSP. Usually System keeps more number of copies with static number. However more referenced files may require for improvising maximum availability.

Redundancy manager calculates the optimal number of copies for the file based on number of references. Numbers of copies are dynamically checked based on changing the number of references.

4. Data Deletion

if Data holder deletes data from Cloud Service Provider, CSP manages the records by removing the duplicated one. If user tries to delete the data from CSP, level of QoS of the corresponding file has been modified which has to be recalculated an optimal number of copies through redundancy manager. if file contains some information other than duplicated data, CSP will not delete the encrypted data but is blocks the data from user. If it is empty, the encrypted data to be removed from CSP.

5. Data Update

When data holder tries to modify an encrypted data, OTP will be generated and provided to the holder based on token generation procedure. After completing the authenticated process, old data M is updated by the User with M' using encryption algorithm and the new encrypted data is replaced in CSP then CSP provides new M' to all the data holders.

Construction of Proposed System Model as follows:

1. User Authentication

- UID and PWD verification
- Send request for Token generation to Private Cloud

Private Cloud

- Generate Token and send to User through OTP
- Private and Public key generated for encryption process through token

2. File Upload Process

- User Verification through OTP [token]
- Duplication Check
    - Check duplicates through Redundancy Manager based on QoS
    - If duplicate found, Create file pointer and store in CSP
    - If not store encrypted file in CSP
    - Redundancy Manager identifies optimal number of copies for corresponding file

3. File Download Process

Public Cloud

- User Verification through OTP [token]
- Decrypt through token and download from Public Cloud

IV. CONCLUSION

Cloud storage services offers on demand virtualization storage resources for improvising storage efficiency. As increasing demand of data store in Cloud, data deduplication is the main technique to improvise storage efficiency. This paper proposed an integrated approach for providing privacy with high end security and improvising storage availability.

REFERENCES

[1] I. Foster, Z. Yong, I. Raicu, and S. Lu, "Cloud Computing and Grid Computing 360- Degree Compared," in Grid Computing Environments Workshop, 2008. GCE '08, 2008, pp. 1-10.

[2] T. Dillon, W. Chen, and E. Chang, "Cloud Computing: Issues and Challenges," in Advanced Information Networking and Applications (AINA), 2010 24th IEEE International Conference on, 2010, pp. 27-33.

[3] T. G. Peter Mell, "The NIST Definition of Cloud Computing," National Institute of Standards and Technology NIST Special Publication 800- 145, September 2011.

[4] SNIA Cloud Storage Initiative, "Implementing, Serving, and Using Cloud Storage,"Whitepaper 2010.

[5] D. Harnik, B. Pinkas, and A. Shulman-Peleg, "Side Channels in Cloud Services: Deduplication in Cloud Storage," Security & Privacy, IEEE,vol. 8, pp. 40-47, 2010.

[6] S. Guo-Zi, D. Yu, C. Dan-Wei, and W. Jie, "Data Backup and Recovery Based on Data De-Duplication," in Artificial Intelligence and Computational Intelligence (AICI), 2010 International Conference on,2010, pp. 379-382.

[7] SNIA, "Advanced Deduplication Concepts," 2011.

[8] V. Javaraiah, "Backup for cloud and disaster recovery for consumers and SMBs," in Advanced Networks and Telecommunication Systems(ANTS), 2011 IEEE 5th International Conference on, 2011, pp. 1-3.

[9] L. L. You, K. T. Pollack, and D. D. E. Long, "Deep Store: An Archival Storage System Architecture," presented at the Proceedings of the 21st International Conference on Data Engineering, 2005.

[10] T. Yujuan, J. Hong, F. Dan, T. Lei, Y. Zhichao, and Z. Guohui, "SAM: A Semantic- Aware Multi-tiered Source De-duplication Framework for Cloud Backup," in Parallel Processing (ICPP), 2010 39th International Conference on, 2010, pp. 614-623.

[11] Waraporn Leesakul , Pentaul Townend and Jie Xu,"Dynamic Data Deduplication in Cloud Storage", 2014 IEEE 8th International Symposium on Service Oriented System Engineering

[12] umedha A Telkar, Dr. M Z Shaikh, "Secured and efficient Cloud Storage Data Deduplication System", International Journal of Advanced Research in Computer and Communication Engineering", Vol 5 issue 1 January 2016.