ISSN:0975-9646

Survey Paper on Deduplicating Data and Secure Auditing in Cloud

Meghana Vijay Kakde, Prof. N.B.Kadu

Department of Computer Engineering,
Pravara Rural College Of Engineering,Loni,Savitribai Phule
Pune University,Pune,India

Abstract:- Data de-duplication is a technique for eliminating duplicate copies of data, and has been widely used in cloud storage to reduce storage space and upload bandwidth. However, there is only one copy for each file stored in cloud even if such a file is owned by a huge number of users. As a result, de-duplication system improves storage utilization while reducing reliability. Furthermore, the challenge of privacy for sensitive data also arises when they are outsourced by users to cloud. Aiming to address the above security challenges, this paper makes the first attempt to formalize the notion of distributed reliable deduplication system. In this paper new distributed deduplication systems with higher reliability in which the data chunks are distributed across multiple cloud servers is being praposed.

Keywords-Reliability,Secret,Sharing,Deduplication, Distributed Storage System.

1. INTRODUCTION

Hiding platform and implementation details unlimited virtualized resources provided to the users as a service is a cloud computing. Presently cloud service provided to the users offered high available storage and massively parallel computing of resources at relatively low costs. But the question is about the cloud users with different privileges store data on cloud is a most challenge issue in managing cloud data storage system.

2. PROBLEM STATEMENT

The problem is to determine how to design secure deduplication systems with higher reliability in cloud computing. Hence it is been proposed in the distributed cloud storage servers into deduplication systems to provide better fault tolerance. To protect data confidentiality, the secret sharing technique is utilized, which is also compatible with the distributed storage systems. To support deduplication, a short cryptographic hash value of the content will also be computed and sent to each storage server as the fingerprint of the fragment stored at each server.

3. LITERATURE SURVEY

The aim is to achieve both data integrity and deduplication in cloud. Hence it is being proposed two secure system as secCloud and secCloud+.Most of the previous deduplication systems have only been considered in a single-server setting. The traditional deduplication methods cannot be directly extended and applied in distributed and multi-server systems. Data network storage overhead by detecting and eliminating redundancy among data. Also, Data reliability is actually a very critical issue in a deduplication storage system because there is only one copy for each file stored in the server shared by all the owners.

4. RELATED WORK

Secure auditing and deduplication is big problem in cloud environment. This technique can be used for securely monitoring server space allocation**.** In this technique how to maintain back up data and remove unwanted file on server.

5. EXISTING SYSTEM

- A number of deduplication systems have been proposed based on various deduplication strategies. such as client-side or server-side deduplications , file-level or block-level deduplication.
- Bellare et al formalized this primitive as message-locked encryption, and explored its application in space efficient secure outsourced storage.
- Li addressed the key-management issue in block-level deduplication by distributing these keys across multiple servers after encrypting the files.
- Bellare et al showed how to protect data confidentiality by transforming the predictable message into unpredictable message.
- The first problem is integrity auditing. The cloud server is able to relieve clients from the heavy burden of storage management and maintenance. The most difference of cloud storage from traditional in-house storage is that the data is transferred via Internet and stored in an uncertain domain ,not under control of the clients at all, which inevitably raises clients great concerns on the integrity of their data.
- The second problem is secure deduplication. The rapid adoption of cloud services is accompanied by increasing volumes of data stored at remote cloud servers. Among these remote stored files, most of them are duplicated: according to a recent survey by EMC, 75% of recent digital data is duplicated copies.
- Unfortunately, this action of deduplication would lead to a number of threats potentially affecting the storage system, for example, a server telling a client that it (i.e., the client) does not need to send the file reveals that some other client has the exact same file, which could be sensitive sometimes. These attacks originate from the reason that the proof that the client owns a given file (or block of data) is solely based on static, short value (in most cases the hash of the file).

Disadvantage

- Data reliability is actually a very critical issue in a deduplication storage system because there is only one copy for each file stored in the server shared by all the owners.
- Most of the previous deduplication systems have only been considered in a single-server setting.
- The traditional deduplication methods cannot be directly extended and applied in distributed and multi-server systems.

6. Proposed System

In this paper, It has shown how to design secure deduplication systems with higher reliability in cloud computing. By introducing the distributed cloud storage servers into deduplication systems to provide better fault tolerance.

To further protect data confidentiality, the secret sharing technique is utilized, which is also compatible with the distributed storage systems. In more details, a file is first split and encoded into fragments by using the technique of secret sharing, instead of encryption mechanisms. These shares will be distributed across multiple independent storage servers.

Furthermore, to support deduplication, a short cryptographic hash value of the content will also be computed and sent to each storage server as the fingerprint of the fragment stored at each server.

Only the data owner who first uploads the data is required to compute and distribute such secret shares, while all following users who own the same data copy do not need to compute and store these shares any more.

To recover data copies, users must access a minimum number of storage servers through authentication and obtain the secret shares to reconstruct the data. In other words, the secret shares of data will only be accessible by the authorized users who own the corresponding data copy.

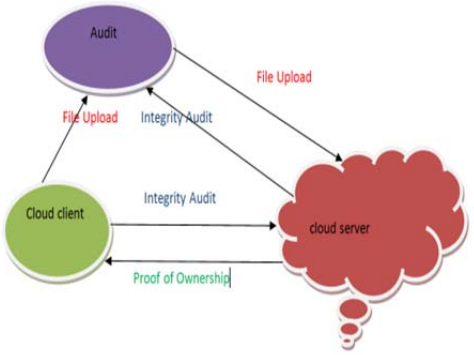
Four new secure deduplication systems are proposed to provide efficient deduplication with high reliability for file-level and block-level deduplication, respectively. The secret splitting technique, instead of traditional encryption methods, is utilized to protect data confidentiality. Specifically, data are split into fragments by using secure secret sharing schemes and stored at different servers.

Advantage

- Distinguishing feature of our proposal is that data integrity, including tag consistency, can be achieved.
- No existing work on secure deduplication can properly address the reliability and tag consistency problem in distributed storage systems.
- The proposed constructions support both file-level and block-level deduplications.
- Security analysis demonstrates that the proposed deduplication systems are secure in terms of the definitions specified in the proposed security model. In more details, confidentiality, reliability and integrity can be achieved in proposed system. Two kinds of collusion attacks are considered in our solutions. These are the collusion attack on the data and the collusion attack against servers. In particular, the data remains secure even if the adversary controls a limited number of storage servers.

This deduplication systems has been implemented using the Ramp secret sharing scheme that enables high reliability and confidentiality levels. The evaluation results demonstrate that the new proposed constructions are efficient and the redundancies are optimized and comparable with the other storage system supporting the same level of reliability.

7. System Architecture



8. Conclusion

It can be conclude that the distributed deduplication systems to improve the reliability of data while achieving the confidentiality of the users' outsourced data without an encryption mechanism. Four constructions were proposed to support file-level and fine-grained block-level data deduplication. The security of tag consistency and integrity were achieved. This deduplication systems has been implemented using the Ramp secret sharing scheme and demonstrated that it incurs small encoding/decoding overhead compared to the network transmission overhead in regular upload/download operations.

References

[1] 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," Communication of the ACM, vol. 53, no. 4, pp.50–58, 2010.

[2] J. Yuan and S. Yu, "Secure and constant cost public cloud storage auditing with deduplication," in IEEE Conference on Communications and Network Security (CNS), 2013, pp. 145–153.

[3] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg, "Proofs of ownership in remote storage systems," in Proceedings of the 18th ACM Conference on Computer and Communications Security. ACM, 2011,pp. 491–500.

[4] S. Keelveedhi, M. Bellare, and T. Ristenpart, "Dupless: Serveraided encryption for deduplicated storage," in Proceedings of the 22Nd USENIX Conference on Security, ser. SEC'13. Washington,D.C.:USENIX Association, 2013, pp. 179–194.

[5] "Message-locked encryption and secure deduplication," in EUROCRYPT, 2013, pp. 296–312.

[6] G. R. Blakley and C. Meadows, "Security of ramp schemes," in Advances in Cryptology: Proceedings of CRYPTO '84, ser. Lecture Notes in Computer Science, G. R. Blakley and D. Chaum, Eds. Springer-Verlag Berlin/Heidelberg, 1985, vol. 196, pp. 242–268.

[7] A. D. Santis and B. Masucci, "Multiple ramp schemes," IEEE Transactions on Information Theory, vol. 45, no. 5, pp. 1720–1728, Jul. 1999.

[8] M. O. Rabin, "Efficient dispersal of information for security, load balancing, and fault tolerance," Journal of the ACM, vol. 36, no. 2, pp. 335–348, Apr. 1989.