INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 3 ISSUE 2 (2017) PAGES 1993 – 1998
RECEIVED: 15-03-2017. PUBLISHED: 04-04-2017.

April 04, 2017

SECURE DISTRIBUTED DEDUPLICATION SYSTEMS WITH IMPROVED RELIABILITY

[1]B.Kathiravan, [2]G.Vadivel,
[1]PG Scholar, Dept Of CSE, Salem college of engineering and technology,
[2]Asst professor, Dept Of CSE, Salem college of engineering and technology.

Abstract

Data deduplication is a method for removing duplicate copies of data, and has been extensively used in cloud storage to decrease storage space and upload bandwidth. On the other hand, there is only one copy for each file stored in cloud even if such a file is owned by a huge number of users. Accordingly, deduplication system progress storage utilization while reducing reliability. In addition, the dare of privacy for sensitive data also take place when they are outsourced by users to cloud. Planning to address the above security test, this paper constructs the first effort to celebrate the idea of scattered reliable deduplication system. This paper recommends a new distributed deduplication systems with upper dependability in which the data chunks are distributed from corner to cornering multiple cloud servers. The safety needs of data privacy and tag stability are also accomplish by introducing a deterministic secret sharing scheme in distributed storage systems, instead of using convergent encryption as in previous deduplication systems.

Keywords : Deduplication, secret sharing, distributed storage system, reliability

1. INTRODUCTION

Cloud computing is comparable to grid computing, a type of computing where unused processing cycles of all computers in a network are harnesses to solve problems too intensive for any stand-alone machine. Cloud computing is an on-demand service that is obtaining mass appeal in corporate data centers. The cloud enables the data center to operate like the Internet and computing resources to be accessed and shared as virtual resources in a secure and scalable manner. Like most technologies, trends start in the enterprise and shift to adoption by small business owners.Deduplication techniques are broadly engaged to backup data and decrease network and storage transparency by notice and eradicate redundancy among data. As an alternative of maintaining multiple data copies with the same content, deduplication reducing redundant data by maintaining only single copy and referring other redundant data to that copy. Deduplication has inward much concentration from both academic world and industry since it can really recover storage utilization and keep storage space, particularly for the applications with high deduplication ratio such as archival storage systems. A number of deduplication systems have been projected based on various deduplication scheme such as client-side or server-side deduplication, file-level or block-level deduplications. Specially, with the advent of cloud storage, data deduplication procedure grow to be more gorgeous and essential for the management of ever-increasing quantity of data in cloud storage services which inspires Endeavour and club to outsource data storage to third-party cloud

providers, If we consider some of the examples as proofs: Cloud storage services, such as, Google Drive, Drop box have been pertaining deduplication to save the network bandwidth and the storage cost with client-side deduplication.

Data consistency is really a very vital issue in a deduplication storage system because there is only one copy for each file accumulates in the server pooled by all the owners. If such a pooled file was lost, a excessively large amount of data becomes unreachable because of the unavailability of all the files that share this file. If the value of a file were calculated in terms of the amount of file data that would be lost in case of behind a single chunk, then the quantity of user data lost when a file in the storage system is spoiled grows with the number of the unity of the chunk. Thus, how to assurance of high data consistency in deduplication system is a vital problem. Most of the preceding deduplication scheme has only been measured in a single-server location. on the other hand, as lots of deduplication systems and cloud storage systems are planned by users and function for higher dependability, particularly in archival storage systems where data are vital and should be potted over long time point. This involve that the deduplication storage systems provide reliability comparable to other high-available systems.

Furthermore, the challenge for data privacy also arises as more and more sensitive data are being outsourced by users to cloud. Encryption mechanisms have usually been utilized to protect the confidentiality before outsourcing data into cloud. Most commercial storage service provider is reluctant to apply encryption over the data because it makes deduplication impossible. The reason is that the traditional encryption mechanisms, including public key encryption and symmetric key encryption, require different users to encrypt their data with their own keys. As a result, identical data copies of different users will lead to different ciphertext. To solve the problems of confidentiality and deduplication, the notion of convergent encryption has been pro-posed and widely adopted to enforce data confidentiality while realizing deduplication. However, these systems achieved confidentiality of outsourced data at the cost of decreased error resilience. Therefore, how to protect both confidentiality and reliability while achieving deduplication in a cloud storage system is still a challenge. Two types of deduplication in terms of the size :(a) block-level deduplication, which find out and eliminate redundancies among data blocks.(b)file-leveldeduplication, which determine redundancies between different files and eradicate these redundancies to decrease ability demands, and The file can be separated into lesser fixed-size. Using fixed-size blocks shorten the calculation of block bound-arise, even as using variable-size blocks.

2. OUR CONTRIBUTIONS

In this paper, we show how to design secure deduplication systems with higher reliability in cloud computing. We introduce 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.

Another distinguishing feature of our proposal is that data integrity, including tag consistency, can be achieved. The traditional deduplication methods cannot be directly extended and applied in distributed and multi-server systems. To explain further, if the same short value is stored at a different cloud storage server to support a duplicate check by using a traditional deduplication method, it cannot resist the collusion attack launched by multiple servers.

In other words, any of the servers can obtain shares of the data stored at the other servers with the same short value as proof of ownership. Furthermore, the tag consistency, which was first formalized by [5] to prevent the duplicate/ciphertext replacement attack, is considered in our protocol. In more details, it prevents a user from uploading a maliciously-generated ciphertext such that its tag is the same with another honestly-generated ciphertext. To achieve this, a deterministic secret sharing method has been formalized and utilized. To our knowledge, no existing work on secure deduplication can properly address the reliability and tag consistency problem in distributed storage systems.

This paper makes the following contributions. Four new secure deduplication systems are pro-posed to provide efficient deduplication with high reliability for file-level and block-level deduplication, respectively. The secret splitting technique, in-stead 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. Our proposed constructions support both file-level and block-level deduplication.
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 our 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.

We implement our deduplication systems using the Ramp secret sharing scheme that enables high re-liability and confidentiality levels. Our 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.

INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 3 ISSUE 2 (2017) PAGES 1993 – 1998
RECEIVED: 15-03-2017. PUBLISHED: 04-04-2017.
April 04, 2017

3. DATA DEDUPLICATION

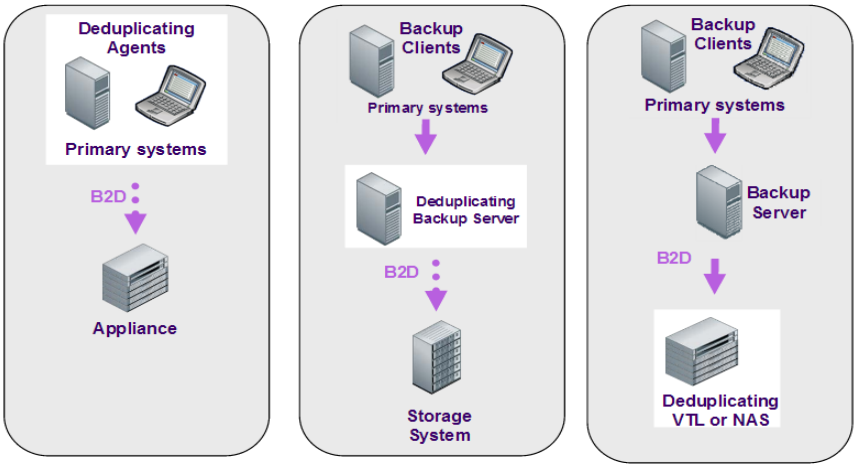


Fig.1. Deduplication for Backup and Recovery

Data deduplication is an advanced technology that can dramatically reduce the amount of backup data stored by eliminating redundant data. Data deduplication maximizes storage utilization while allowing IT to retain more near line backup data for a longer time. This tremendously improves the efficiency of disk based backup, changing the way data is protected. In general, data deduplication compares new data with existing data from previous backup or archiving jobs, and eliminates the redundancies. Advantages include improved storage efficiency and cost savings, as well as bandwidth minimization for less expensive and faster offsite replication of backup data.

4. THE DISTRIBUTED DEDUPLICATION SYSTEMS

The distributed deduplication systems future aim is to reliably store data in the cloud while achieving privacy and consistency. Its main objective is to allow deduplication and distributed storage of the data diagonally multiple storage servers. As an alternative encrypting the data to keep the privacy of the data, new structures put on the top-secret intense technique to split data into shards. These shards will then be distributed transversely in multiple storage servers.

Message authentication code

A message authentication code (MAC) is a short piece of information used to authenticate a message and to provide integrity and authenticity assurances on the message. In our construction, the message authentication code is applied to achieve the integrity of the outsourced stored files. It can be easily constructed with a keyed (cryptographic) hash function, which takes input as a secret key and an arbitrary-length file that needs to be authenticated, and outputs a MAC. Only users with the same key generating the MAC can verify the correctness of the MAC value and detect whether the file has been changed or not.

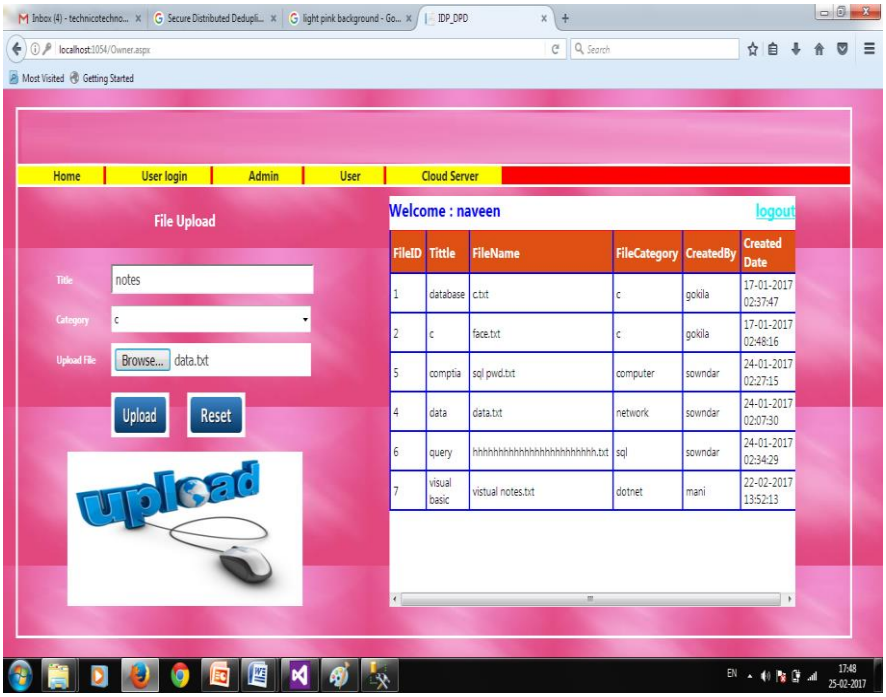
INTERNATIONAL RESEARCH JOURNAL IN ADVANCED ENGINEERING AND TECHNOLOGY (IRJAET) E - ISSN: 2454-4752 P - ISSN : 2454-4744
VOL 3 ISSUE 2 (2017) PAGES 1993 – 1998
RECEIVED: 15-03-2017. PUBLISHED: 04-04-2017.
April 04, 2017

File-Level Distributed Deduplication System

To support efficient duplicate check, tags for each file will be computed and are sent to S-CSPs. To prevent a collusion attack launched by the S-CSPs, the tags stored at different storage servers are computationally independent and different.

Block-Level Distributed Deduplication System

The fine-grained block-level distributed deduplication. In a block-level deduplication system, the user also needs to first perform the file-level deduplication before uploading his file. If no duplicate is found, the user divides this file into blocks and performs block-level deduplication. The system setup is the same as the file-level deduplication system, except the block size parameter will be defined additionally. Next, we give the details of the algorithms of File Upload and File Download.



CONCLUSION

The proposed distributed deduplication systems are to increase the consistency of data however attaining the privacy of the user's outsourced data without an encryption appliance. The security of tag consistency and integrity were attained. The implementation of deduplication systems using the Ramp secret sharing scheme here gives the demonstration that it acquires small encoding/decoding overhead compared to the network transmission overhead in regular download /upload operations.

REFERENCE

[1] M. O. Rabin, "Fingerprinting by random polynomials,"Center forRes. Comput. Technol., Harvard Univ., Tech. Rep. TR-CSE-03-01, 1981.

[2] J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer, "Reclaiming space from duplicate files in a serverless distributed file system," in Proc. 22nd Int. Conf. Distrib. Comput. Syst., 2002, pp. 617–624.

[3] M. Bellare, S. Keelveedhi, and T. Ristenpart, "Dupless: Serveraided encryption for deduplicated storage," in Proc. 22nd USENIX Conf. Secur. Symp., 2005, pp. 179–194.

[4] M. Bellare, S. Keelveedhi, and T. Ristenpart," Message-locked encryption and secure deduplication," in Proc. EUROCRYPT, 2008, pp. 296–312.

[5] G. R. Blakley and C. Meadows, "Security of ramp schemes," in Proc. Adv. Cryptol., 2008, vol. 196, pp. 242–268.

[6] A. D. Santis and B. Masucci, "Multiple ramp schemes," IEEE Trans. Inf. Theory, vol. 45, no. 5, pp. 1720–1728, Jul. 2009.