A SURVEY ON SECURE DE-DUPLICATION APPROACH IN STORAGE CLOUD SERVICE PROVIDER WITH EFFICIENT AND RELIABLE CONVERGENT KEY MANAGEMENT

Miss. Vaishali Bhusari
ME Student,
Dept. Computer Science & Engg.
G.H. Raisoni College of Engineering & Management
Amravati, M.P.

Prof. Avinash Wade
Head of Department CSE
G.H. Raisoni College of Engineering & Management
Amrawati, M.P.

ABSTRACT: Data deduplication is a technique for eliminating duplicate copies of data, and has been widely used in cloud storage to reduce storage space and upload bandwidth. Promising as it is, an arising challenge is to perform secure deduplication in cloud storage. Although convergent encryption has been extensively adopted for secure deduplication, a critical issue of making convergent encryption practical is to efficiently and reliably manage a huge number of convergent keys. Data deduplication looks for redundancy of sequences of bytes across very large comparison windows. Sequences of data (over 8 KB long) are compared to the history of other such sequences and it is ideal for highly redundant operations like backup which requires repeatedly copying and storing the same data set multiple times for recovery purpose. To protect the confidentiality of sensitive data while supporting deduplication, convergent encryption technique has been designed to encrypt the data. Convergent encryption enables duplicate files to coalesce into the space of a single file, even the files are encrypted with different user's keys.

KEYWORDS*:* RACS, *Convergent encryption, FADE,* Vanish, PoW

INTRODUCTION

Cloud computing provides unlimited "virtualized" resources to users as services across the whole Internet, while hiding platform and implementation details. Today's CSP(cloud service providers) offer both highly available storage and especially parallel computing resources at relatively low costs. As cloud computing becomes ubiquitous, an amount of data is being stored and shared by users with specified privilege he cloud in t, which define the access rights of the stored data. One significant challenge of cloud storage services is the management of the ever-increasing volume of data. Cloud storage provides cost effective resource usage as a service to users. Every user has large amount of data to store it in a secured storage area. To make data management scalable in cloud storage data deduplication technique is used. Data deduplication is the process is identifying duplicate data stored in the cloud, it will reduce the cost and maximize the storage space to the user. Deduplication can take place in file-level and block-level. Data deduplication brings a lot of benefits. Security and privacy concerns arise as users sensitive data are outsourced to cloud storage. In traditional encryption, data confidentiality and security is incompatible with data deduplication. Specifically, traditional encryption requires different users to encrypt their data with their own keys. Thus, same data copies of different users will lead to different ciphertexts, it makes deduplication impossible.To provide secure deduplication, and convergent encryption has been proposed to achieve data confidentiality. It encrypts/decrypts a data copy with a convergent key, which is derived from the data using the cryptographic hash functions to obtain the same ciphertext. To prevent unauthorized access to the data secure proof of ownership protocol is used to prove that user owns the same data and Differential authorization duplicate check is used to allow only a authorized user to check duplicates of data that is for each user set of privileges is issued during the system initialization.

LITERATURE SURVEY

To observe that there is a good deal of sharing between the data on typical laptops. For example, most (but not all) of the system files are likely to be shared with at least one other user. But equally importantly, it would. significantly reduce the time required for backups in most cases – upgrading an operating system, or downloading a new music file should not require any additional backup time at all if someone else has already backed-up those same files. There has been a lot of interest recently in de duplication techniques, using content-addressable storage (CAS). This is designed to address exactly the above problem. However, most of these solutions are intended for use in a local file system or SAN. This has two major drawbacks:

(1) Clients must send the data to the remote file system before the duplication is detected – this forfeits the potential saving in network traffic and time.

(2) Any encryption occurs on the server, hence exposing sensitive information to the owner of the service – this is usually not appropriate for many of the files on a typical laptop which are essentially personal rather than corporated. Backing up to cloud-based storage becomes increasing popular in recent years. The main benefits of using cloud storage are lower server maintenance cost, cheaper long term operational cost, and sometimes enhanced data safety via a vendor's own geographically diverse data replication. In particular, the benefits of employing a cloud-based secondary storage are:

1. New cloud services can be added easily on the backup server to provide enhanced data safety and to reduce the risk of vendor lock-in.

2. Upload cost to cloud storage can be reduced via data aggregation techniques. In ASE (Asymmetric searchable encryption) schemes are appropriate in any setting where the party searching over the data is different from the party that generates it. The main advantage of ASE is functionality while the main disadvantages are inefficiency and weaker.

EXISTING SYSTEM

The cloud environment is a large open distributed system. It is important to preserve the data, as well as privacy of the users. Existing techniques in cloud storage are Redundant Array of Cloud Storage (RACS): This RACS is used to stored data over multiple vendors. This is act as a proxy and performs the operation between the client and server. This method is simple and easy to work with. The drawback is single proxy cloud easily become bottleneck. Secure Overlay Service (SOS): SOS is mainly used to solve the distributed denial of service, the idea provide by this solution is very complicated. It is unclear about the optimal solution. Vanish: This technique of storage is all data become unreliable after some specific period of time for the security purpose. The data will be deleted after the particular time period with the knowledge of the owner who created the data. This technique is more expensive and it requires large Distributed Hash Table (DHT). FADE: This FADE is secure overlay in cloud storage with assured deletion. The data owner can be sure of the deleted file. Only the deleted part of the file is considered not the accessing data. This method is more complex to implement. The most critical challenge of cloud storage is the increasing volume of data. So, to overcome such a problem data deduplication techniques has been used. The deduplication techniques identify redundant copy of data and it will not store the redundant data on the cloud storage instead it will keep only the physical data and refer to the redundant to that stored copy. Thus such techniques improve the storage utilization of the cloud storage.

DIFFERENT APPROCHES

The most critical challenge of cloud storage is the increasing volume of data. So, to overcome such a problem data deduplication techniques has been used. The deduplication techniques identify redundant copy of data and it will not store the redundant data on the cloud storage instead it will keep only the physical data and refer to the redundant to that stored copy. Thus such techniques improve the storage utilization of the cloud storage. Although the data deduplication gives more benefits, security and privacy concerns arise because the user's sensitive data is susceptible to both the outsider and insider attacks. So, considering the traditional encryption techniques to secure the users sensitive data. Traditional encryption provides data confidentiality but it is incompatible with deduplication. As in traditional encryption different users encrypt their data with their own keys. Thus, the identical data of the different users will lead to different ciphertext which is making the data deduplication impossible. A virtual private storage service which is based on the cryptographic techniques, which provides secure cloud storage. It provides confidentiality, integrity and non- repudiation to the data stored on cloud storage. A new system which provides secure and efficient access to outsourced data. Here the end user sends a request for data access to the data owner, after that the data owner will send back an encryption key and access certificate to an end user, and then the end user will send that access certificate to the data storage provider and the data storage provider will send the encrypted data blocks to the end user.

4.1. Traditional Encryption

The data deduplication gives more benefits, security and privacy concerns arise because the user's sensitive data is susceptible to both the outsider and insider attacks. So, considering the traditional encryption techniques to secure the users sensitive data. Traditional encryption provides data confidentiality but it is incompatible with deduplication. As in traditional encryption different users encrypt their data with their own keys. Thus, the identical data of the different users will lead to different ciphertext which is making the data deduplication impossible. A virtual private storage service which is based on the cryptographic techniques, which provides secure cloud storage. It provides confidentiality, integrity and non- repudiation to the data stored on cloud storage. A new system which provides secure and efficient access to outsourced data. Here the end user sends a request for data access to the data owner, after that the data owner will send back an encryption key and access certificate to an end user, and then the end user will send that access certificate to the data storage provider and the data storage provider will send the encrypted data blocks to the end user.

Advantages:

1) It has a low storage overhead.

Disadvantages:

1) It requires support from the cloud side.
2) Multiple policies combination is not provided.

System supports the use of multiple policies. Here we focus on new approach which is named as FADE. Authors in their study have proposed a new protocol called vanish which provides data privacy and self-deleting data. The study is mainly focused on the data and it could be able to access for a limited period of time. After the time expiry, the data is not accessible to the users nor to the data owner. Vanish protocol is applicable for only sensitive data. To has self-deleting property, the activities takes place are, Vanish first encrypts user's data locally by taking the help of encryption key and the encryption key will not be known to the user also, then it destroy local copies of key and after that it sprinkles bits in DHT randomly.

Disadvantages:

1) The drawback of this system is it provides the assured deletion based on time. Even the legitimate users may not be able to access the data after time expiration.

Disadvantage of all Traditional Encryption

Approaches:

1) Incompatible with data deduplication.

PROPOSED TECHNIQUE

5.1 Convergent Encryption

Convergent encryption ensures data isolation in deduplication. This primitive as message locked encryption and explored its application in space efficient secure outsourced storage .The problem and showed a secure convergent encryption for efficient encryption without considering issues of the key management and block level deduplication. To encrypt a file using convergent encryption, a client computes a cryptographically strong hash of the file content. The file is then encrypted using this hash value as a key. The hash value is then encrypted using the public keys of all authorized readers of the file and these encrypted values are attached to the file as metadata. Convergent encryption enables identical encrypted files to be recognized as identical, but there remains the problem of performing this identification across a large number of machines in a robust and decentralized manner .There are also more than a few implementations of convergent implementations of different convergent encryption variants for secure deduplication. It is known that some commercial cloud storage providers also deploy convergent encryption.

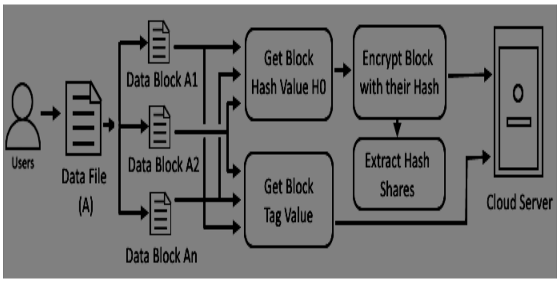

Fig 1: Convergent Encryption

5.2 Proof of ownership

The PoW ("proofs of ownership") for deduplication systems such that a client can efficiently confirm to the cloud storage server that he/she owns a file without uploading the file itself. Some PoW constructions based on the system proposed to enable client-side deduplication which includes the bounded leakage setting. The system proposed another efficient PoW scheme by choosing the projection of a file onto some randomly selected bit positions as the file confirmation. Note that all the above schemes do not consider data privacy. Recently, PoW for encrypted files, but they do not address how to reduce the key management overhead. Proof-of-ownership is a protocol in two parts between two players on a joint input F (which is the input file). First the verifier summarizes to itself the input file F and generates a (shorter) verification information v, the prover and verifier engage in an interactive protocol in which the prover has F and the verifier only has v, at the end of which the verifier either accepts or rejects. Hence a proof of-ownership is specified by a summary function S(_) (which could randomized and takes the input file F and a security parameter) and an interactive two-party protocol _(P $ V).

5.3Ramp Secret Sharing

Dekey uses the Ramp secret sharing scheme (RSSS) to store convergent keys. Specifically, the (n,k,r)-RSSS (where n > k > r > 0) generates n shares from a secret such that 1) the secret can be recovered from any k shares but cannot be recovered from fewer than k shares, and 2) no information about the secret can be deduced from any r shares. It is known that when r ¼ 0, the (n,k,r)-RSSS becomes the (n,k,r) Rabin's Information Dispersal Algorithm (IDA), when r ¼ k À 1, the (n,k,r)-RSSS becomes the (n,k,r)Shamir's Secret Sharing Scheme (SSSS) . The (n,k,r)-RSSS builds on two primitive functions:

1) Share divides a secret S into (k-r) pieces of equal size, generates r random pieces of the same size, and encodes the k pieces using a non-systematic k-of-n erasure code 1 into n shares of the same size.

2) Recover takes any k out of n shares as inputs and then outputs the original secret S.

Dekey uses RSSS to provide a tunable key management mechanism to balance among confidentiality, reliability, storage overhead, and performance.

SYSTEM MODEL

A data outsourcing model used by Dekey. There are three entities, namely: the user, the storage cloud service provider (S-CSP), and the key management cloud service provider (KM-CSP), as elaborated below.

1) User: A user is an entity that wants to outsource data storage to the S-CSP and access the data later. To save the upload bandwidth, the user only uploads unique data but does not upload any duplicate data, which may be owned by the same user or different users.

2) S-CSP: The S-CSP provides the data outsourcing service and stores data on behalf of the users. To reduce the storage cost, the S-CSP eliminates the storage of redundant data via deduplication and keeps only unique data.

3) KM-CSP: A KM-CSP maintains convergent keys for users, and provides users with minimal storage and computation services to facilitate key management. For fault tolerance of key management, we consider a quorum of KM-CSPs, each being an independent entity. Each convergent key is distributed across multiple KM-CSPs using RSSS.

A data copy to be either a whole file or a smaller-size block, and this leads to two types of deduplication:

1) file-level deduplication, which eliminates the storage of any redundant files, and

2) block-level deduplication, which divides a file into smaller fixed-size or variable-size blocks and eliminates the storage of any redundant blocks. Using fixed-size blocks simplifies the computations of block boundaries, while using variable- size blocks (e.g., based on Rabin fingerprinting) provides better deduplication efficiency. We deploy our deduplication mechanism in both file and block levels. Specifically, to upload a file, a user first performs the file- level duplicate check. If the file is a duplicate, then all its blocks must be duplicates as well; otherwise, the user further performs the block-level duplicate check and identifies the unique blocks to be uploaded. Each data copy (i.e., a file or a block) is associated with a tag for the duplicate check. All data copies and tags will be stored in the S-CSP.

CONCLUSION

We propose an efficient and reliable convergent encryption scheme for secure deduplication. RSSS applies deduplication among convergent keys and distributes convergent key shares across multiple key servers, while preserving semantic security of convergent keys and confidentiality of outsourced data. We also explained detailed about secure file upload and downloading process.

REFERENCES

1) P. Anderson and L. Zhang. Fast and secure laptop backups with encrypted de-duplication. USENIX LISA, 2010.
2) M. Bellare, S. Keelveedhi, and T. Ristenpart. Dupless: Serveraided encryption for deduplicated storage. In USENIX Security Symposium, 2013.
3) M. Bellare, S. Keelveedhi, T. Ristenpart. Message locked encryption and secure deduplication. In EUROCRYPT, pages 296–312, 2013.
4) 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. ICDCS, 2002, pp. 617-624.
5) M.W. Storer, K. Greenan, D.D.E. Long, and E.L. Miller, “Secure Data Deduplication,” in Proc. StorageSS, 2008, pp. 1-10.
6) A. Yun, C. Shi, and Y. Kim, “On Protecting Integrity and Confidentiality of Cryptographic File System for Outsourced Storage,” in Proc. ACM CCSW, Nov. 2009, pp. 67-76.
7) A. Rahumed, H.C.H. Chen, Y. Tang, P.P.C. Lee, and J.C.S. Lui, “A secure Cloud Backup System with Assured Deletion and Version Control,” in Proc. 3rd Int’l Workshop Security Cloud Comput., 2011, pp. 160-167.