International Journal of Interdisciplinary Innovative Research &Development (IJIIRD)
ISSN: 2456-236X
Vol. 05 Special Issue 01 | 2020

Data Security by Using Convergent Keys

Kalyani R. Gawande[1], R. R. Bhure[2]

[1] M.E Student, Dept. of Computer Engineering, P.R.Pote Amravati,India
[2] Ph.D. Scholar, Dept. of Computer Engineering, P.R.Pote Amravati,India

ABSTRACT

Secure deduplication is a technique to remove duplicate copies of storage data, and provides security to them. To reduce storage space and upload bandwidth in cloud storage deduplication has been a well-known technique. For that purpose convergent encryption has been extensively adopt for secure deduplication, critical issue of making convergent encryption practical is to efficiently and reliably manage a huge number of convergent keys. The basic idea in this, is that eliminate duplicate copies of storage data and limit the damage of stolen data if decrease the value of that stolen information to the attacker. Here it makes the first attempt to formally address the problem of achieving efficient and reliable key management in secure deduplication. Here, it first introduce a baseline approach in which each user holds an independent master key for encrypting the convergent keys and outsourcing them. However, such a baseline key management scheme generates an enormous number of keys with the increasing number of users and requires users to dedicatedly protect the master keys. To this end, here propose Dekey technology. Dekey new construction in which users do not need to manage any keys on their own but instead securely distribute the convergent key shares across multiple servers for insider attacker. As a proof of concept, implementation of Dekey by using the Ramp secret sharing scheme and demonstrate that Dekey incurs limited overhead in realistic environment.

Keywords: Deduplication, Storage Space, Convergent encryption key management, Dekey, Ramp Secret Sharing.

1. INTRODUCTION

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. One critical challenge of today's cloud storage services is the management of the ever-increasing volume of data. To make data management scalable deduplication we are use convergent Encryption for secure deduplication services.

Businesses, especially start-ups, small and medium businesses (SMBs), are increasingly opting for outsourcing data and computation to the Cloud. Today's commercial cloud storage services, such as Dropbox, Mozy, and Memopal, have been applying deduplication to user data to save maintenance cost[1].From a user's point of view, data outsourcing raises security and privacy concerns. However, deduplication, while improving storage and bandwidth efficiency, is compatible with Convergent key management. Specifically, traditional encryption requires different users to encrypt their data with their own keys. Many proposals have been made to secure remote data in the Cloud using encryption and standard access controls. It is fair to say all of the standard approaches have been demonstrated to fail from time to time for a variety of reasons, including insider attacks, mis-configured services, faulty implementations, buggy code, and the creative construction of effective and sophisticated attacks not envisioned by the implementers of security procedures. Building a trustworthy cloud computing environment is not enough, because accidents continue to happen, and when they do, and information gets lost, there is no way to get it back. One needs to prepare for such accidents.

The basic idea is that,limit the damage of stolen data, if we decrease the value of that stolen information from the attacker. Achieve this through a 'preventive' disinformation attack. Secure deduplication services can be implemented given two additional security features.

2. RELATED WORK

Most of the existing data for each user stored in the cloud and the demand of long term continuous assurance of their data safety, the problem of verifying correctness of data storage in the cloud becomes even more challenging. Cloud Computing is not just a third party data warehouse. The data stored in the cloud may be frequently updated by the users, including insertion, deletion, modification, appending, reordering, etc. One critical challenge of today's cloud storage services is the management of the ever-increasing volume of data. According to the analysis report of IDC, the volume of data in the wild is expected to reach 40 trillion gigabytes in 2020.

3. EXISTING SYSTEM

In the hybrid cloud architecture it consist of a public cloud and a private cloud. Unlike existing data deduplication systems, the private cloud is involved as a proxy to allow data owner/users to securely perform duplicate check with differential privileges[7]. Such architecture is practical and has attracted much attention from researchers. The data owners only outsource their data storage by utilizing public cloud while the data operation is managed in private cloud. The user is only allowed to perform the duplicate check for files marked with the corresponding privileges. Furthermore, it enhance the system in security. Specifically, it present an advanced scheme to support stronger security by encrypting the file with differential privilege keys. In this way, the users without corresponding privileges cannot perform the duplicate check for deduplication to protect the confidentiality of sensitive data while supporting deduplication, the convergent encryption technique has been used to encrypt the data before outsourcing. To better protect data security, this paper makes the first attempt to formally address the problem of authorized data deduplication. Different from traditional deduplication systems, the differential privileges of users are further considered in duplicate check besides the data itself. Also present several new deduplication constructions supporting authorized duplicate check in a hybrid cloud architecture. Security analysis demonstrates that the scheme is secure in terms of the definitions specified in the proposed security model. As a proof of concept, it implement a prototype of our proposed authorized duplicate check scheme and conduct test bed experiments using prototype. Showing the authorized duplicate check scheme incurs minimal overhead compared to normal operations.

3.1 Limitations of Existing System:

The baseline approach suffers two critical deployment issues. First, it is inefficient, as it will generate an enormous number of keys with the increasing number of users. Specifically, each user must associate an encrypted convergent key with each block of its outsourced encrypted data copies, so as to later restore the data copies. Although different users may share the same data copies, they must have their own set of convergent keys so that no other users can access their files. Second, the baseline approach is unreliable, as it requires each user to dedicatedly protect his own master key. If the master key is accidentally lost, then the user data cannot be recovered; if it is compromised by attackers, then the user data will be leaked.

4. BASELINE APPROACH

Subject to chunking/blocking method, two types of deduplication method/strategies are there:

4.1 File level chunking

This algorithm, examines whole file as one chunk and does not split the file into small blocks Hence in this method, for the whole file only one index value is generated and this index value, is further compared with previously saved index values. As there is only one index value for every file, there would be relatively lesser entries in the index table. This concept would decrease the total storage size and in the same index table more entries for unique indexes can be made as well. This file level segregation fails when there is a slight change in file data, because it will generate index for complete file again rather, it should generate index for only changed data which in turns decreases the ratio of deduplication elimination and throughput of the system.

4.2 Block Level Chunking

It is of two types as given below:

4.2.1 Fixed-Size Chunking: Using this technique the data file is further partitioned into fixed sized blocks or chunks. Fixed size chunking solves the problem which had arisen in file level chunking as in this method the index value is generated for different blocks instead of files.
Therefore when any data is changed index of only that block is changed not of the whole file.[14]. On other hand many small chunks are created for large files resulting in consumption of more storage space for large number of index value and metadata.

4.2.2 Variable-Size Chunking: Using this technique, the data file is partitioned in numerous small sized blocks or chunks which are rather of variables size than being of same fixed size and the file is segregated on the basis of the content of the data than same fixed size value. Therefore solving the problem and eliminating the drawbacks of the fixed size chunking. It is to be noted that in fixed sized chunking data boundaries do not change even when there is a change in the data whereas in later i.e. variable sized chunking data boundaries are of variable size depending upon the different parameters and even these boundaries can be shifted when there is any change in file some deletion of data or file occurs. Therefore when there is any change in file, fewer boundaries are needed to be changed.

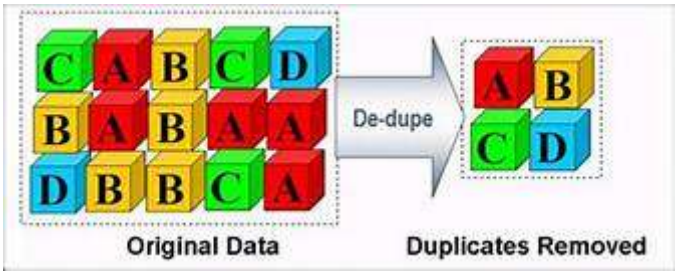

Fig : 4.1 Deduplication Process

5. PROPOSED SYSTEM

Dekey new construction in which users do not need to manage any keys on their own but instead securely distribute the convergent key shares across multiple servers for insider attacker. As a proof of concept, implementation of Dekey by using the Ramp secret sharing scheme and demonstrate that Dekey incurs limited overhead in realistic environments.

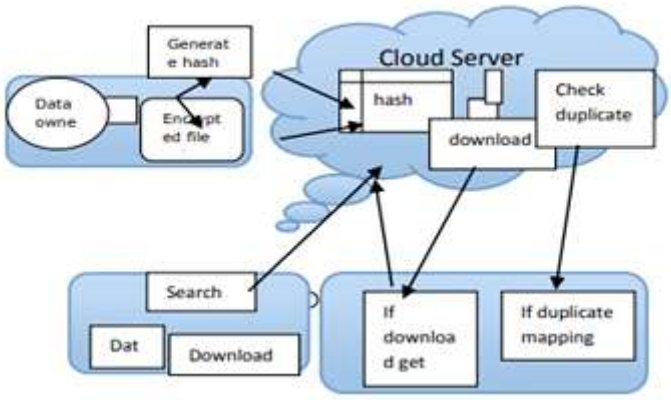

Fig : 5.1 System Architecture of Secured deduplication using Dekey concept

A new construction called Dekey, which offers efficiency and reliability agreements for convergent key management on both consumer and cloud storage space regions. Our design is to be relevant deduplication to the convergent keys and force secret sharing techniques. Specially, we create secret distributes for the convergent keys and share them transversely several independent key servers. Merely the first consumer who uploads the data is necessary to calculate and share such secret shares, while all following consumers who own the identical data copy need not compute and store these shares over again. To recuperate data copies, a consumer should access a least amount of key servers through authentication and attain the secret shares to recreate the convergent keys. In other words, the secret distributes of a convergent key will merely be accessible by the authoritative consumers who own the corresponding data copy. This considerably decreases the storage overhead of the convergent keys and creates the key management reliable next to failures and attacks. In addition, the project also judges the revocation of consumers in the given group [9]. If the original consumer of the group intimates the server with a consumer's (B) revocation, then the server rejects the proof of ownership submitted by that consumer (B).

5.1 Ramp Secret Sharing Scheme:

Dekey make use of the Ramp secret sharing scheme (RSSS) [13, 14] to store convergent keys. Specially, the RSSS creates n distributes from a secret such the secret can be improved from any k distributes but cannot be improved from less than k distributes, and no data about the secret can be presumed from any r distributes. It is identified that when r = 0, the (n,K,0)-RSSS grow to be the (n,k) Rabin's Information Dispersal Algorithm (IDA) when r =k-1, the (n,k,k-1)-RSSS grow to be the (n,k) Shamir's Secret Sharing Scheme (SSSS).

6. APPLICATIONS

• Data deduplication is that it reduces the amount of disk or tape that organizations need to buy, which in turn reduces costs.
• NetApp reports that in some cases, deduplication can reduce storage requirements up to 95 percent, but the type of data you're trying to deduplicate and the amount of file sharing your organization does will influence your own deduplication ratio.

• While deduplication can be applied to data stored on tape, the relatively high costs of disk storage make deduplication a very popular option for disk-based systems.
• Eliminating extra copies of data saves money not only on direct disk hardware costs, but also on related costs, like electricity, cooling, maintenance, floor space, etc.
• Deduplication can also reduce the amount of network bandwidth required for backup processes, and in some cases, it can speed up the backup and recovery process.

7. ADVANTAGES

• The detection of masquerade activity.
• The confusion of the attacker and the additional costs incurred to distinguish real from fake information.
• The deterrence effect which, although hard to measure, plays a significant role in preventing masquerade activity by risk-averse attackers.

8. CONCLUSIONS

DeKey an efficient and reliable convergent key management scheme for secure deduplication. DeKey applies deduplication among convergent keys and distributes convergent key shares across multiple key servers, while reserving semantic security of convergent keys and confidentiality of outsourced data

9. REFERENCES

[1]M. Bellare, A. Desai, E. Jokipii, and P. Rogaway. A Concrete Security Treatment of Symmetric Encryption: Analysis of the DES Modes of Operation. Proceedings of the 38th Symposium on Foundations of Computer Science, IEEE, 1997.
[2] Ayushi "A Symmetric Key Cryptographic Algorithm " International Journal of Computer Applications (0975 - 8887) ©2010 Volume 1 – No. 15
[3] Abdul Wahid Soomro, Nizamuddin, Arif Iqbal Umar, Noorul Amin." Secured Symmetric Key Cryptographic Algorithm for Small Amount of Data" 3rd International Conference on Computer & Emerging Technologies (ICCET 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] W. J. Bolosky, J. R. Douceur, D. Ely, and M. Theimer, "Feasibility of a Serverless Distributed File System Deployed on an Existing Set of Desktop PCs", SIGMETRICS 2000, ACM, 2000, pp. 34-43.
[6] A. Adya, W. J. Bolosky, M. Castro, R. Chaiken, G. Cermak, J. R. Douceur, J. Howell, J. R. Lorch, M. Theimer, and R. Wattenhofer. FARSITE: Federated, available, and reliable storage for an incompletely trusted environment. In Proceedings of the 5th Symposium on Operating Systems Design and Implementation (OSDI), Boston, MA, Dec.2002. USENIX.
[7] R. Anderson and E. Biham, "Two Practical and Provably Secure Block Ciphers: BEAR and LION", 3rd International Workshop on Fast Software Encryption, 1996, pp. 113-120.