Indian Journal of Science and Technology, *Vol 9(4), DOI: 10.17485/ijst/2016/v9i4/87036, January 2016*
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

A Survey on Data Redundancy Check in a Hybrid Cloud by using Convergent Encryption

N. R. Anitha Rani*, S. K. Ram Kumar and P. Prem Kumar

CSE Department, KLNCE, Pottapalayam, Madurai - 625001, Tamil Nadu, India; anitharani.in@gmail.com, ramkumar880@gmail.com, premkumar20041971@gmail.com

Abstract

Background/Objectives: Data deduplication is the elimination of redundant data within an existing environment. It is mainly used in cloud storage environment for bandwidth optimization and to enhance storage space. This paper focuses on different dedupe techniques with a secure cloud. **Methods/Statistical Analysis:** Symmetric key encryption is used to prevent unauthorized user from accessing the data. Convergent key encryption algorithm is used to encrypt and maintain data confidentiality. **Findings:** In the proposed system the authorized data deduplication is achieved with minimal overhead when compared with traditional deduplication.

Keywords: Convergent Key, Cloud Storage, Data Confidentiality, Data Deduplication, Hybrid Cloud

1. Introduction

In the current IT environment every organization is in need to invest time and budget to scale up the IT infrastructure such as hardware, software and services. Cloud computing provides computing over the internet. Cloud services consist of highly optimized virtual Data Center that provides various software, hardware and information resource to the needful users. The organization can simply connect to the cloud and use the available resource on pay per user basis. This helps the company to avoid capital expenditure on additional on-premise infrastructure resources and instantly scale up or scale down according to business requirements.

Data deduplication shown in Figure 1 is essential in today's world to achieve bandwidth optimization and to lower the storage space by eliminating redundant data copies.

Data deduplication can be performed at two levels:

- File Level Deduplication: Eliminates duplicate copies of identical files[1].
- Block Level Deduplication: Eliminates duplicate data blocks in non identical files[2].

The first and foremost challenge in the current IT world is to prevent sensitive data from being accessed by unauthorized users. In conventional encryption different users use their own key to encrypt data and as a result different cipher text is generated for the same file which belongs to different users. This makes deduplication highly impossible. To overcome this symmetric key encryption is used to generate same key to encrypt and decrypt identical files of different users. Convergent key encryption provides data confidentiality and checks for redundant copies of data. Secure proof of ownership

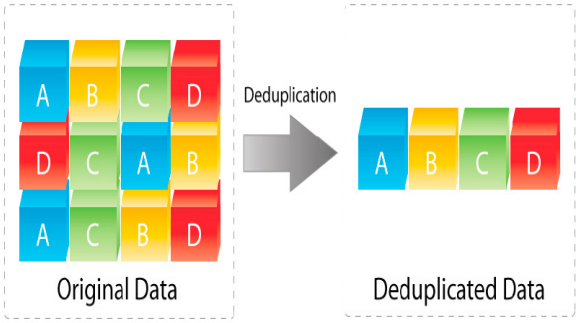

Figure 1. Data deduplication.

*Author for correspondence

Table 1. Comparison of different methods of data deduplication

Author	Title	Features	Result
Jin Li, Xiaofeng Chen, Mingqiang Li, Jingwei Li, Patrick P.C. Lee, and Wenjing Lou[3]	Secure Deduplication with Efficient and Reliable Convergent Key Management[3]	• Focuses on deduplication of convergent key. • Supports both file level and block level deduplication. • Reliable key management.	• Convergent key shared across multiple servers to make data deduplication process to function effectively.
Li, J.; Chen, X.; Huang, X.; Tang, S.; Xiang, Y.; Hassan, M.; Alelaiwi, A[5]	Secure Distributed Deduplication Systems with Improved Reliability[5]	• Minimize network and storage overhead by detecting and eliminating redundancy among data blocks. • Ensures fault tolerance even if some nodes fails.	• Enhance data efficiency with minimal overhead
Jingwei Li, Xie, D.,Cai, Z[6]	Secure Auditing and Deduplicating Data in Cloud[6]	• Checks the integrity through auditing by clustering files. • Enhances data security.	• An effective cloud service management is done by enforcing integrity auditing and reduce computation work by the user during upload and auditing phase.
Mihir Bellare, Sriram Keelveedhi, Thomas Ristenpart[7]	DupLESS: Server-aided encryption for deduplicated storage[7]	• Maintain data confidentiality and data integrity • Enhance bandwidth optimization	• Efficient storage interface which incur less overhead on resources.

prevents unauthorized users to access the data. Different methods of data deduplication are shown in Table 1.

1.1 A Secure Deduplication with Efficient and Reliable Convergent Key Management[3]

The elimination of redundant data within an existing environment is called Data deduplication. It is mainly used in cloud storage environment to reduce both bandwidth usage and storage space. The proposed system aims at achieving proficient and trustworthy convergent key management through convergent key deduplication and to reduce enormous number of key generation with increase in number of users[4].

1.2 Methodology used in this System

- Convergent Encryption: Ensures protection against unauthorized users in deduplication.
- Dekey: It creates secret shares on original convergent key and sends the shares across multiple key management cloud service providers. The multiple users who share the same block can access the same corresponding convergent keys.

1.3 Characteristics

- Increase the storage in cloud by applying deduplication on convergent key and ensures reliable key management.
- File Level and Block Level Deduplication is supported by Dekey.

2. Secure Distributed Deduplication Systems with Improved Reliability[5]

Data deduplication is the process of eliminating redundant copies of data in cloud storage. The main aim of data deduplication is to reduce storage space and upload bandwidth. When the data is outsourced there arises confrontation regarding privacy of sensitive data. In the proposed system the data chunks are distributed across multiple cloud servers with higher reliability and

uses deterministic secret sharing scheme in distributed storage for data confidentiality.

2.1 Methodology used in this System

- Secret Sharing Scheme: It consists of two algorithms Share and Recover
- Share: The secret key is divided and shared
- Recover: With enough shares the secret key can be extracted and recovered.
- Tag Generation Algorithm: Maps the original data copy and the tag which is used by the user to perform duplicate check with the server.
- Proof of Ownership: Takes inputs as file and an index then generates a tag as output which is used for proof of ownership.
- Message Authentication Code: It is a short piece of information used to authenticate message and provides integrity.

2.2 Characteristics

- Minimize network and storage overhead by detecting and eliminating redundancy among data blocks.
- Ensures fault tolerance even if certain number of nodes fail.

3. Secure Auditing and Deduplicating Data in Cloud[6]

In outsourced cloud storage, data maintenance is not highly reliable. The main aim of this project is to resolve the problem of integrity auditing and enhance deduplication along with data reliability in the outsourced cloud storage. In this system two secure systems, are constructed (i.e.) SecCloud and SecCloud+, which helps to achieve both data integrity and secure deduplication.

3.1 Methodology used in this System

3.1.1 SecCloud

SecCloud initiates an auditing entity with a preservation of a MapReduce cloud. It generates a set of data tags and then sends these tags while uploading the files to the cloud server by the client. Further it also audits the integrity of data stored in the cloud. The SecCloud system supports file level deduplication and intercept the leakage of side channel information.

3.1.2 SecCloud+

SecCloud+ system grants integrity auditing and data deduplication on encrypted files. The client with the key, generates convergent key to encrypt the file before uploading to the cloud server. This system permits the guarantee of file confidentiality. The convergent key to encrypt the uploading file is generated and controlled by a secret key. It supports file level deduplication and sector level deduplication.

3.2 Characteristics

- It provides integrity auditing by clustering the files by removing duplicates.
- The duplicate files are detected with a single copy of the file by mapping data with the existing file in the cloud.

4. Dupless Server - Aided Encryption for Deduplicated Storage[7]

Cloud Storage Service providers perform deduplication to save storage space only by storing unique data (by avoiding redundant copies) by using Message Locked Encryption technique. In the proposed system the clients encrypt the data by message - based keys acquired from a key- server through PRF (pseudorandom function) protocol. The client's stores encrypted data with an existing service and perform deduplication.

4.1 Methodology used in this System

- Message locked encryption: It is a cryptographic framework method used to deduplication of cipher text. The encryption is done using keys derived from the message itself.

4.2 Characteristics

- Enhances the Data Confidentiality and Data Integrity through message locked encryption.
- Ensures bandwidth optimization through data deduplication.

5. Methods used in Secure Deduplication in a Hybrid Cloud[8] (Figure 2)

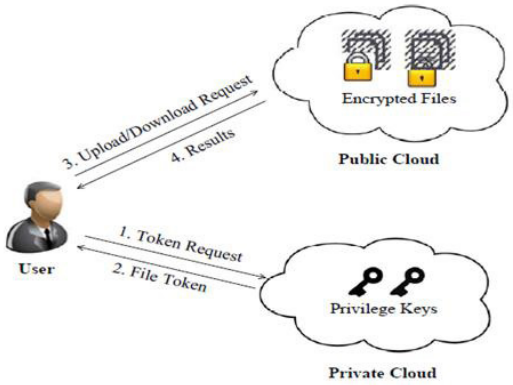

Figure 2. System architecture.

5.1 Symmetric Encryption

Encryption is an important part of computing factor in today's environment. Encryption is a process of taking data and encoding it into a form that cannot be read by un authorized users. Symmetric key[9] uses the common secret key to encrypt and decrypt data. It is much faster than public key encryption. When using symmetric key encryption[10] following measures have to be followed.

- Key needs to be stored securely.
- If another user wants to decrypt the encrypted file then a secure channel is required to transfer the key.

5.2 Convergent Encryption

Convergent Encryption provides only authorized users to access the data (data confidentiality) in data deduplication. Convergent Encryption uses hash functions like SHA to generate cipher text. The hash value for files is generated first and if both the files are same then they produce same cipher text. The convergent key encryption uses this hash value as a key to encrypt the file and also derives a tag to detect duplicates. When two files are identical then generated tags and their cipher text is also same. The user first passes the tag to the server side to check if same copies of data are already stored.

5.3 Proof of Ownership

It is an interactive algorithm[11] run by the user and storage server and it act as an interface between the user and the storage server. It enables the users to prove their rights on the data to the storage server.

5.4 Identification Protocol

Identification protocol[12] consists of two phase: Proof and Verify. In proof phase the user provides his identity to the verifier by performing identification proof related to his identity. In the verify phase the verifier checks proof and accepts only valid proof and rejects the invalid proof.

6. Conclusion

In this paper we have surveyed the various data deduplication techniques and also highlighted the features of various systems. The secure authorized data deduplication system provides a secure data deduplication on cloud storage. This system goes one step ahead by providing differential privileges to different users and also enhances security by introducing hybrid cloud approach.

7. References

1. Li J, Li YK, Chen X, Lou PPC. A hybrid cloud approach for secure authorized deduplication. IEEE Transactions on Parallel and Distributed Systems. 2015; 26(5):1206-16.
2. Elmagarmid AK, Ipeirotis PG, Verykios PG. Duplicate record detection: A survey. IEEE Transactions on Knowledge and Data Engineering. 2007 Jan; 19(1):1-16.
3. Li J, Chen X, Li M, Li J, Lee PPC. Secure deduplication with efficient and reliable convergent key management. IEEE Transactions on Parallel and Distributed Systems. 2014; 25(6):15-25.
4. Lin J, Chen X, Xhafa F, Barolli L. Secure deduplication storage systems supporting keyword search. ACM Journal of Computer and System Sciences. 2015; 81(8):1532-41.
5. Li J, Chen X, Huang X, Tang S, Xiang Y, Hassan M, Tang S. Secure distributed deduplication systems with improved reliability. IEEE Transactions on Computers. 2015; 64(12):569-79.
6. Li J, Xie D, Cai Z. Secure auditing and deduplicating data in cloud. IEEE Transactions on Computers. 2015; (99):1.
7. Bellare M, Keelveedhi S, Ristenpart T. DupLESS: Server-aided encryption for deduplicated storage. Proceedings of the 22nd USENIX Conference on Security; 2013. p. 179-94.
8. Bugiel S, Nurnberger S, Sadeghi A-R, Schneider T. Twin clouds: An architecture for secure cloud computing. International Conference on Communications and Multimedia Security; 2011. p. 1-11 .

9. Chaoling L, Yue C , Yanzhou Z. A data assured deletion scheme in cloud storage. IEEE Transactions on China Communications. 2014; 11(4):98-110.
10. Sugumar R, Sheik Imam SB. Symmetric encryption algorithm to secure outsourced data in public cloud storage. Indian Journal of Science and Technology. 2015 Sep; 8(23):1-5.
11. Halevi S, Harnik D, Pinkas B, Shulman-Peleg A. Proofs of ownership in remote storage systems. Proceedings of the 18th ACM Conference on Computer and Communications Security; 2011. p. 491-500.
12. Bellare M, Namprempre C, Neven G. Security proofs for identity-based identification and signature schemes. Journal of Cryptology-ACM. 2008; 22(1):1-61.