

# A Survey on Deduplication check method and self-destruction for cloud computing security.

Asha Kanade, Aishwarya Dhumal, Vaishnavi  
Chavan, Snehal Khandale, Prof P.V. Kharat.

kanadeasha7896@gmail.com, aishwaryadhupal8@gmail.com,  
khandalesnehal96@gmail.com, vaishnavichavan182@gmail.com,  
kharat.pramila@gmail.com

NESGOI, Faculty of Engineering, Naigaon-Pune

Pune University, Pune, Maharashtra, India

**Abstract:** In cloud computing, most of the communication is done in consideration of file processing, and therefore it turns out to be very critical and important to deliver competent method for data security. The data stored on the cloud is private data of the user so it must not be tampered by other entities. The backup of the data will be taken timely into the backup directory. The main focus is on data deduplication to provide security services over cloud for stored files. Data deduplication is data compression that is used to eliminate the matching copies of data. This procedure is frequently used for reducing the storage space and save bandwidth over cloud. Novel deduplication structures are offered for supporting authorized duplicate check. In this survey paper we have discussed a system to enhance the security; the data uploaded by a user is shuffled between the number of directories within cloud after a particular interval of time to avoid the tracking of the data.

**Keywords:** - Deduplication, storage system, Duplicate Data Elimination (DDE), Data encryption, Hybrid cloud, Security.

## I. Introduction

Cloud computing is the most popular technology nowa day it is used by most of the social media sites and organizations to store the data. Data stored in clouds is highly sensitive, for example Social networking records and medical records. Providing privacy and security on that data is major concern.

Data deduplication checking is one of the repeated data elimination method by keeping unique copy in storage system to reduce the wastage of storage space on cloud. This method will help to improve the efficiency and searches are faster. Deduplication

identifies the files which are already available on cloud storage and it uses the "reference pointer" to find out unnecessary chunks; this is also known as the storage capacity optimization. Data deduplication has two types as file level de-deduplication and block level data de-duplication. Instead of keeping multiple copies of same files like image, text, video files de-duplication checks and maintain unique copy. Deduplication will work on following four conditions:

1. Same file name with same contents.
2. Same file name with different contents.
3. Different file name with same contents.
4. Different file name with different contents.

The data stored on the cloud is sensitive data of the user so there can be chances of data hacking by unauthorized entities. We propose a system to improve the security, and when user upload files that can rotate on different directories on timely basis so data tracking can be avoided. The replica of the data will be stored on virtual cloud. The proposed system gives the deduplication checking and also provides security with easy use of the cloud.

Hashing is the method which is used for content matching of two different files. For calculating hash values, tag of the file (which we want to upload) will be generated.

On the basis of architecture cloud services are divided into three fundamental models. These three fundamental models are often referred to as the SPI model where SPI stand for Software as a service, Platform as a service and Infrastructure as a service.

## II. Literature Survey

There are two important requirements for cloud storage which are as follows:

- 1] Data integrity: -A] Proof of Retrievability B] Proof of Data Possession
- 2] storage efficiency: -A] Proof of Ownership Yuan et al. [1] The system reduces the wastage of storage space and protect the data sensitivity. Bellare et al. [2] Increase the data sensitivity by sending the predictable message. They use new method in the hybrid Cloud structure, to provide better data security. J. Li [3]. In secure Deduplication they find problem to achieving effective and consistent key management. Then they use decryption key to enable key management to gain different reliability and confidentiality levels X. Chen[4] Message-Locked Encryption is a method to get secure deduplication. A Message-Locked Encryption symmetric method is used for plain to cipher text and cipher to plain text of message. M. Bellare[5]. Brute-force attacks for duplication uses Dupless architecture which provides protected storage. By using PRF protocol clients saves the data with an

available service, however achieve strong privacy M. Bellare[6].

## III. Proposed System

The diagrammatic representation of our system is as shown in figure. The main motive behind data deduplication is to provide security to the cloud. We propose a system to improve the security, and when user upload files that can rotate on different directories on timely basis so data tracking can be avoided. The replica of the data will be stored on virtual cloud. The proposed system gives the deduplication checking and also provides security with easy use of the cloud.

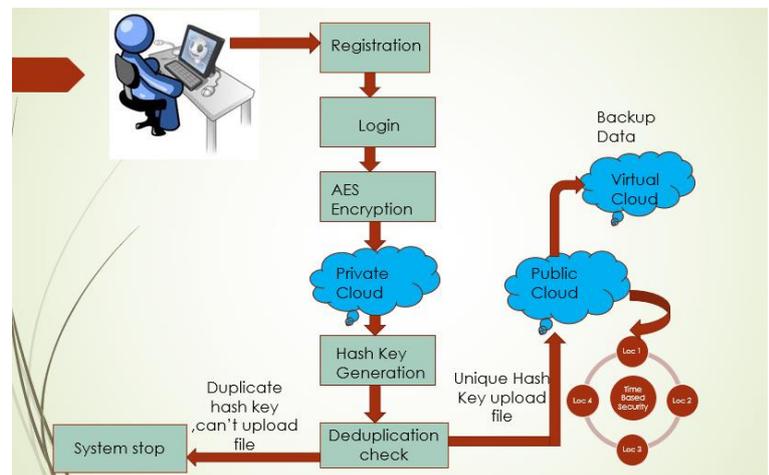


Figure 1: The proposed System Architecture.

## IV. System Features

### 4.1 Functional Requirements

- System should be avoiding data duplication and reduce wastage of space on cloud.
- System should be maintaining data theft.
- System should reduce the data tracking.
- System should provide virtual cloud for backup of a data.

## 4.2 External Interface Requirements

### 4.2.1 User Interfaces

- Registration Form
- Login Form
- Minimum 10 GB Hard disk.
- Core 2 dual or above Software.

### 4.2.3 Software Interfaces:

- Operating System: Windows, Ubuntu
- Eclipse Kepler
- Apache Tomcat Server 7
- JDK 1.7
- MySQL Workbench 5.0

## V. Technical Specification

### 5.1 Advantages

- Reduce storage space.
- Provide high security to sensitive data.
- Take virtual backup of data.

### 5.2 Disadvantages

- Need of internet connection is compulsory to check the data deduplication over cloud.

### 5.3 Applications

- This application can be used by government to check duplicate files on their cloud storage, maintain security, and take backup of highly sensitive data.
- This application can be used by all type of industries, organizations, peoples to store their data efficiently, securely on cloud.

## VI. Conclusion

The main motive of data deduplication is to eliminate repeated data and keeps original physical copy of data. Data deduplication used to avoid wastage of cloud storage and

- File Upload Form
- Deduplication check Form
- Data backup request form

### 4.2.2 Hardware Interfaces

- 4 GB RAM or More.
- improves efficiency and saves the network bandwidth.

## Acknowledgment

Special thanks to the in Charge Prof P. V. Kharat for her guidance and constant supervision as well as for giving helpful information related to the project work and also backing for project completion. We would like to show our special thankfulness to the professional person for giving us such attention and time.

## References

- [1] M. Bellare, S. Keelveedhi, and T. Ristenpart. Dupless: Serveraided encryption for deduplicated storage. In USENIX Security Symposium, 2013.
- [2] J. Yuan and S. Yu. Secure and constant cost public cloud storage auditing with deduplication. IACR Cryptology ePrint Archive, 2013:149, 2013.
- [3] J. Li, Y. Kit Li, X. Chen, P. P. C. Lee and W. Lou, "A Hybrid Cloud Approach for Secure Authorized Deduplication", In IEEE Transactions on Parallel and Distributed Systems, DOI:10.1109/TPDS.2014.2318320, (2015) April 7, pp. 1206-1216.
- [4] J. Li, X. Chen, M. Li, J. Li and P. P. C. Lee, "Secure Deduplication with Efficient and Reliable Convergent Key Management", In IEEE Transactions on Parallel and Distributed Systems, DOI: 10.1109/TPDS.2013.284, (2014) May 12, pp. 1615-1625

- [5] M. Bellare, S. Keelveedhi and T. Ristenpart, "Message-locked encryption and securededuplication", Proceedings of EUROCRYPT, Athens Greece, (2013) March 3, pp. 296-312.
- [6] M. Bellare, S. Keelveedhi and T. Ristenpart, "Dupless: Serveraided encryption for deduplicated storage", In USENIX Security Symposium, Washington DC, (2013) August 14-16, pp. 179-194.
- [7] V. Satish Radia and D. Kumar Singh, "Secure Deduplication Techniques: A Study", International Journal of Computer Applications, (2016), pp. 41-43.
- [8] S. Bugiel, S. Nurnberger, A. Sadeghi, and T. Schneider. Twin clouds: An architecture for secure cloud computing. In Workshop on Cryptography and Security in Clouds (WCSC 2011), 2011.
- [9] K. Zhang, X. Zhou, Y. Chen, X. Wang, and Y. Ruan. Sedic: privacyaware data intensive computing on hybrid clouds. In Proceedings of the 18th ACM conference on Computer and communications security, CCS'11, pages 515–526, New York, NY, USA, 2011. ACM.
- [10] R. S. Sandhu, E. J. Coyne, H. L. Feinstein, and C. E. Youman. Role-based access control models. IEEE Computer, 29:38–47, Feb 1996.
- [11] J. Stanek, A. Sorniotti, E. Androulaki, and L. Kencl. A secure data deduplication scheme for cloud storage. In Technical Report, 2013.
- [12] M. W. Storer, K. Greenan, D. D. E. Long, and E. L. Miller. Secure data deduplication. In Proc. of StorageSS, 2008.
- [13] Z. Wilcox-O'Hearn and B. Warner. Tahoe: the least-authority filesystem. In Proc. of ACM StorageSS, 2008.
- [14] J. Xu, E.-C. Chang, and J. Zhou. Weak leakage-resilient client-side deduplication of encrypted data in cloud storage. In ASIACCS, pages 195–206, 2013.
- [15] J. Yuan and S. Yu. Secure and constant cost public cloud storage auditing with deduplication. IACR Cryptology ePrint Archive, 2013:149, 2013.