

# Authorized Deduplication: An Approach for Secure Cloud Environment

<sup>1</sup> Bachi Reddy Sarath Kumar Reddy, <sup>2</sup> G.Viswanath

<sup>1</sup> (M.Tech) –CSE, ST. Johns College of Engineering & Technology,  
Yemmiganur, Kurnool (Dist)-Andhra Pradesh India.

<sup>2</sup> Assistant Professor, Dept. of CSE, ST. Johns College of Engineering & Technology,  
Yemmiganur, Kurnool (Dist)-Andhra Pradesh India.

**Abstract:** - nowadays a huge volume of data generated and stored into cloud data base from four corners of the universe, Data deduplication turns out to be more need for Cloud storage providers. By putting away a distinct duplicate copy of data, cloud Data extraordinarily diminishes their capacity and Data exchange costs. The benefits of data deduplication accompany a high cost regarding new security and protection contests. We propose secure data deduplication mechanism, a safe and efficient Storage service which guarantees bit level secure data deduplication and Data classified ness in the meantime. To perform secure access scheming user may satisfy access privileges issued by data owner at cloud level towards access restricting from unauthorized users or adversaries.

**Keywords:-** Deduplication, authorized duplicate check, confidentiality, File level Check, Block Level Check, Convergent key, Metadata Supervisor.

## 1. Introduction

Cloud computing provides seemingly unlimited “virtualized” resources to users as services across the entire Web while concealing stage and usage points of interest. With the possibly unending storage room offered by cloud suppliers, clients tend to use as much space as they can, and vendors always search for strategies meant to minimize repetitive information and amplify space investment funds. A system which has been broadly received is cross-client deduplication. The straightforward thought behind deduplication is to store copy information (either records or pieces) just once. Along these lines, if a client needs to transfer a document (square) which is as of now put away, the cloud supplier will add the customer to the proprietor rundown of that record (piece). Deduplication has demonstrated to accomplish high space and expense investment funds, and numerous distributed storage suppliers are right now embracing it. Deduplication can decrease stockpiling needs by up to 90-95% for reinforcement applications [11] and up to 68% in standard

record frameworks [23]. Alongside low possession require the insurance of their information and secrecy ensures through encryption. Lamentably, deduplication and encryption are two clashing innovations. While the point of deduplication is to recognize indistinguishable information portions and store them just once, the aftereffect of encryption is to make two indistinguishable information sections indistinct in the wake of being scrambled. This implies that if information is encoded by clients in a standard manner, the distributed storage supplier cannot make a difference deduplication since two indistinguishable information portions will be distinctive after Encryption. Then again, if data is not scrambled by Information proprietors, confidentiality cannot be ensured, and the information is not secured against inquisitive distributed storage suppliers. A procedure which has been proposed to meet these two clashing pre requisites is Convergent encryption whereby the encryption key is usually the consequence of the hash of the information portion. Albeit Convergent encryption is by all accounts a decent possibility to accomplish confidentiality and deduplication

in the meantime, it sadly experiences different surely understood shortcomings [15], [24] word reference assaults: an aggressor why should capable figure or foresee a document can without much of a stretch infer the potential encryption key and check whether the record is officially put away at the distributed storage supplier or not. In this paper, we adapt to the inborn security exposures of focalized encryption and propose secure information deduplication instrument, which safeguards the consolidated focal points of deduplication and Convergent encryption.

Data owner can restrict unauthorized access rights by performing fine-grained (Privileges) access controlling scheme where data owner defined set of access attribute sets before outsourcing to public cloud, if any user wants to access that file user need to satisfy the data owner access attribute sets, if it's matched the data owner allows him to access that data by sending set of access privileges. Data deduplication will be done in the secure manner by proving proof of the ownership.

## 2. Related Work

In this section, we review some related works concerned with security and privacy issues in the cloud. Also, we discuss the work which adopts similar techniques as our approach but serves different purposes.

### 2.1. Security and Privacy issues in the cloud:

Only the authorized persons need to access the data from the cloud. To ensure the integrity of user authentication, lack of security mechanism which will keep track usage of information in the cloud? As with all cloud computing security challenges, it is the responsibility of the user to ensure that the cloud provider has taken all necessary security measures to protect the user's data and the access to that data.

Deduplication is the technique that is most effective most widely used, but when it is applied across the multiple users, the cross-user deduplication tends to have too many serious privacy implications. Simple mechanisms can be used which can enable the cross-user deduplication which will reduce the risks of the data leakage.

In previous deduplication systems cannot support differential authorization duplicate check, which is important in many applications. In such an authorized deduplication system, each user has issued a set of privileges during system initialization. The overview of the cloud deduplication is as follow:

### 2.2. Deduplication

According to the data granularity, deduplication strategies can be categorized into two broad categories: file-level deduplication [29] and block-level deduplication [17], which is nowadays the most common strategy. In block-level deduplication, the block size can either be fixed or variable [27]. Another categorization criterion is the location at which deduplication is performed: if data are deduplicated at the client, then it is called source-based deduplication, otherwise target-based. In source-based deduplication, the client first hashes each data segment he wishes to upload and sends these results to the storage provider to check whether such data are already stored: thus only "un-deduplicated" data segments will be uploaded by the user. While deduplication at the client side can achieve bandwidth savings, it, unfortunately, can make the system vulnerable to side-channel attacks [19] whereby attackers can immediately discover whether a certain data is stored or not. On the other hand, by deduplicating data at the storage provider, the system is protected against side-channel attacks, but such solution does not decrease the communication overhead.

### 2.3. Convergent Key Encryption

The basic idea of convergent key encryption (CKE) is to derive the encryption key from the hash of the plaintext. The simplest implementation of convergent encryption can be defined as follows: Data owner derives the encryption key from his/her message  $M$  such that  $K = H(M)$ , where  $H$  is a cryptographic hash function; Data owner can encrypt the message with this key, hence:  $C = E(K, M) = E(H(M), M)$ , where  $E$  is a block cipher. By applying this technique, two users with two identical plaintexts will obtain two identical ciphertexts since the encryption key is the same; hence the cloud storage provider will be able to perform deduplication on such ciphertexts. Furthermore, encryption keys are generated, retained and protected by users. As the encryption key is deterministically generated from the plaintext, users do not have to interact with each other for establishing an agreement on the key to encrypt a given plaintext. Therefore, convergent encryption seems to be a good candidate for the adoption of encryption and deduplication in the cloud storage domain.

## 3. System Study

### 3.1. Presented System:

In our presented system, data deduplication performed at service provider level without considering user privileges, data get stored at cloud server level with related privileges

keys. More over there is a lack of security while accessing from cloud servers due to weak access controlling schemes like coarse-grained approach was performed at the client level.

There might be possibilities are there to access the data by adversaries. If data duplication occurs at block level i.e., if the context of the file is same or File level, i.e., name of the archive is same then duplication functioning will be executed, in order to function data deduplication mechanism system has verified POW (Proof of the ownership), and then check the label tags which are maintained by the cloud service provider.

### Disadvantages:

- Lack of user privacy
- Lack of data confidentiality
- Lack of data integrity
- Unsecured data duplication mechanism performed
- Redundant data avoidance systems cannot support differential authorization duplicate check

## 3.2. The Proposed System

The idea of data deduplication with secured manner is the foremost objective of the proposed system, in this connection we proposed secure data deduplication mechanism by distinguishing sensitive and non-sensitive data at data uploading into cloud level and apply the crypto algorithm for sensitive data by applying this data get secured and authorized

## 4. System Implementations

**User:** User must be registered for to upload data into clouds by providing required info... like name, password, and email, mobile.

**Data Owner:** - Data owner will make account in our application by using the registration form and by using the

### 4.1. Restrict From Unauthorized Access

When the user wants to access the data from the public cloud, that user needs to authorized by the data owner by having privilege keys taken from data owner through any one of the secured communication systems, i.e., E-mail, unless and until get the access key from the data owner. Authorization system does not allow any access rights to words protecting access from unauthorized users.

### 4.2. Algorithm Used

his/her user name and password he can login in to our application they can upload and download data from our cloud server the data will provide security by encrypting the data in the files and giving privileges to other data users according to user requests and that given privileges information will be sent to users registered e-mail.

**Data deduplication with secure manner:** while data uploading by the user into the public cloud, the identification of duplicate data will be notified by showing the warning pop msg to users if the user wants to upload existing file again, still user want to upload a file the new file needs to update with the existing file. While user uploading data into public cloud user can distinguish sensitive and non-sensitive data and can provide encryption for only sensitive data. If any unauthorized user wants to access or the user did not have particular privileges (as read write, if user is having read privileges, but they want to access file (downloading like that)) immediately message alert need to send to for a particular data owner

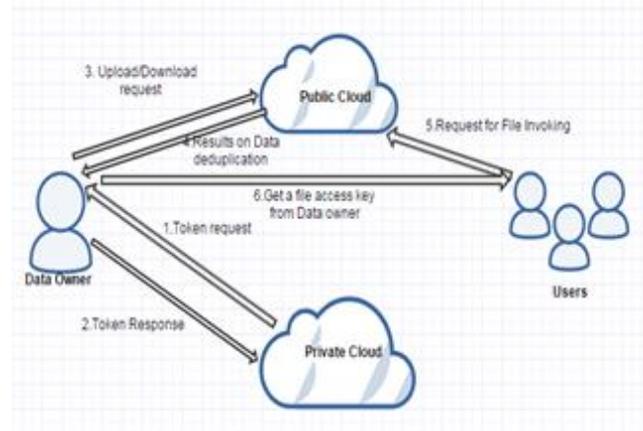


Figure 1. System Architecture

Here in this section to provide secure data accessing from the public cloud, while uploading the data into public cloud by the data owner, data need to be encrypted using secure cryptographic and convergent key encryption algorithm. It is a symmetric cryptographic algorithm, which performs secured data encryption and decryption by using the same key, which leads easy key management along with high performance. In this concern, encrypted data will be protected from cloud provider as well adversaries.

## 5. Conclusion

In this paper we addressed secure data deduplication process for every uploaded data into public cloud by separate the process of sensitive data and non-sensitive data, while accessing data from public cloud only authorized users can access the data for the sake of data read/write. In the interests of data privacy from public cloud or attackers, we can encrypt data owner will give only sensitive data and privileges.

## 6. Further Enhancements:

1. Generating One Time Password while user login for to upload data into the cloud. OTP will generate and send to the user registered mobile number.

### 6.1. Algorithm Used To Generate OTP:

#### HMAC-based One-time Password Algorithm

- K be a secret key
- C be a counter
  - $HMAC(K,C) = SHA1(K \oplus 0x5c5c... \square SHA1(K \oplus 0x3636... \square C))$  be an HMAC calculated with the SHA-1 cryptographic hash algorithm
  - Truncate be a function that selects 4 bytes from the result of the HMAC in a defined manner Then HOTP(K,C) is mathematically defined by  $HOTP(K,C) = Truncate(HMAC(K,C)) \& 0x7FFFFFFF$

The mask 0x7FFFFFFF sets the result's most significant bit to zero. This avoids problems if the result is interpreted as a signed number as some processors do.[1]

For HOTP to be useful for an individual to input to a system, the result must be converted into a HOTP value, a 6–8 digits number that is implementation dependent.

$HOTP-Value = HOTP(K, C) \bmod 10^d$ , where d is the desired number of digits

HOTP can be used to authenticate a user in a system via an authentication server. Also, if some more steps are carried out (the server calculates subsequent OTP value and sends/displays it to the user who checks it against subsequent OTP value calculated by his token), the user can also authenticate the validation server.

### 6.2. Elimination of Identical Data

Hear in the Present system we eliminate duplicate copies of data by file name, now in future, we can add some extra flavors to the current system for to remove identical copies of data efficiently by using pattern matching algorithm. Where this algorithm detects multiple copies of identical picks and shows to the user whether the user is still want to upload all identical copies of data or not.

## 7. References

- [1] P. Anderson and L. Zhang. "Fast and secure laptop backups with encrypted de-duplication". In Proc. of USENIX LISA, 2010.
- [2] M. Bellare, S. Keelveedhi, and T. Ristenpart. "Dupless: Server aided encryption for deduplicated storage". In USENIX Security Symposium, 2013.
- [3] Pasquale Puzio, Refik Molva ,MelekOnen,"CloudDedup: Secure Deduplication with Encrypted Data for Cloud Storage", SecludIT and EURECOM, France.
- [4] Iuon –Chang Lin, Po-ching Chien ,”Data Deduplication Scheme for Cloud Storage” International Journal of Computer and Control(IJ3C),Vol1,No.2(2012)
- [5] Shai Halevi, Danny Harnik, Benny Pinkas,”Proof of Ownership in Remote Storage System”, IBM T.J.Watson Research Center, IBM Haifa Research Lab, Bar Ilan University,2011.
- [6] M. Shyamala Devi, V.Vimal Khanna,Naveen Balaji ”Enhanced Dynamic Whole File De-Duplication(DWFD) for Space Optimization in Private Cloud Storage Backup”,IACSIT, August,2014.
- [7] Weak Leakage-Resilient Client –Side deduplication of Encrypted Data in Cloud Storage” Institute for Info Comm Research,Singapore,2013
- [8] Tanupriya Chaudhari , Himanshu shrivastav, Vasudha Vashisht, ”A Secure Decentralized Cloud Computing Environment over Peer to Peer”,IJCSMC, April,2013
- [9] Mihir Bellare, Sriram keelveedhi,Thomas Ristenart ,”DupLESS: Server Aided Encryption for Deduplicated storage” University of California, San Diego2013.
- [10] Luna SA HSM. <http://bit.ly/17CDPm1>.
- [11] Opendedup. <http://opendedup.org/>.
- [12] Atul Adya, William J Bolosky, Miguel Castro, Gerald Cermak, Ronnie Chaiken, John R Douceur, Jon Howell, Jacob R Lorch, Marvin Theimer, and Roger P Wattenhofer. Farsite: Federated, available, and reliable storage for an incompletely trusted environment. ACM SIGOPS Operating Systems Review, 36(SI):1–14, 2002.

- [13] Mihir Bellare, Alexandra Boldyreva, and Adam O'Neill. Deterministic and efficiently searchable encryption. In *Advances in Cryptology-CRYPTO 2007*, pages 535–552. Springer, 2007.
- [14] Mihir Bellare, Sriram Keelveedhi, and Thomas Ristenpart. Dupless: Server-aided encryption for deduplicated storage. 2013.
- [15] Mihir Bellare, Sriram Keelveedhi, and Thomas Ristenpart. Message-locked encryption and secure deduplication. In *Advances in Cryptology–EUROCRYPT 2013*, pages 296–312. Springer, 2013.
- [16] Kevin D. Bowers, Ari Juels, and Alina Oprea. Hail: a high-availability and integrity layer for cloud storage. In *Proceedings of the 16th ACM conference on Computer and communications security, CCS '09*, pages 187–198, New York, NY, USA, 2009. ACM.
- [17] Landon P Cox, Christopher D Murray, and Brian D Noble. Pastiche: Making backup cheap and easy. *ACM SIGOPS Operating Systems Review*, 36(SI):285–298, 2002.
- [18] John R Douceur, Atul Adya, William J Bolosky, P Simon, and Marvin Theimer. Reclaiming space from duplicate files in a serverless distributed file system. In *Distributed Computing Systems, 2002. Proceedings. 22nd International Conference on*, pages 617–624. IEEE, 2002.
- [19] Danny Harnik, Benny Pinkas, and Alexandra Shulman-Peleg. Side channels in cloud services: Deduplication in cloud storage. *Security & Privacy, IEEE*, 8(6):40–47, 2010.
- [20] Jollu Jayachandrudu, M. Sri lakshmi, Dr. S. Prem Kumar, "Enhanced Independent Access to Encrypted Cloud Databases," *International Journal of Computer Engineering In Research Trends*, vol.2, no.9, pp. 589-593, 2015.
- [21] K. Naga Maha Lakshmi, A. Shiva Kumar, "Secure Data Deduplication and Data accessing among Multi-cloud Framework," *International Journal of Computer Engineering In Research Trends*, vol.2, no.10, pp. 687-693, 2015.
- [22] D. Jayanarayana Reddy, M. Janardhan, U. Veeresh, "Secure Data Deduplication over Distributed Cloud Server Framework with Effective User Revocation and Load Balancing," *International Journal of Computer Engineering In Research Trends*, vol.4, no.2, pp. 57-62, 2017.
- [23] Aradhyula Venkata Ramu, D. Ravi Kiran, "Novel Approach for Secure Data deduplication System." *International Journal of Computer Engineering In Research Trends*, vol.2, no.10, pp. 896-901, 2015.
- [24] K. Sudhamani, P. Rama Rao, R. Vara Prasad, "Secure Auditing and Deduplicating Data in Cloud," *International Journal of Computer Engineering In Research Trends*, vol.3, no.1, pp. 1-5, 2016.