IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710


# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

## IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Special Issue 2, April 2023

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

|| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | SJIF: 8.423|| Volume 12, Special Issue 2, April 2023 ||

8th National Conference on Advanced Computing Technologies (NCACT'23)

Organized by

Department of CSE, Velammal Engineering College, Chennnai, India

# Secure and Efficient Data Privacy with Deduplication in Cloud Environment

**Mrs. S. Daya Florance[1], Yuvan Kumar.S[2], Vinoth Kumar.M[2], Ratish.G[2], Vijay Narayanen.K[2]**

Assistant Professor, Department of Computer Engineering, Velammal Engineering College, Chennai,

Tamil Nadu, India[1]

U.G. Scholar, Department of Computer Engineering, Velammal Engineering College, Chennai,

Tamil Nadu, India[2]

**ABSTRACT:**Information deduplication procedure has been broadly received by commercial cloud capacity suppliers, which is both critical and fundamental in adapting with the hazardous development of information.To assist ensure the security of clients touchy information within the outsourced capacity mode, numerous secure information deduplication plans have been planned and connected in different scenarios.Among these plans, secure and proficient re- encryption for scrambled information deduplication pulled in the consideration of numerous researchers, and numerous arrangements have been planned to bolster energetic proprietorship administration.In this paper, we center on the re-encryption deduplication capacity framework and appear that the as of late planned lightweight rekeyingaware scrambled deduplication plot (REED) is powerless to an assault

**KEYWORDS:** Security, Cloud service Providers(CSP), Shared key, Deduplication

## I. INTRODUCTION

Cloud technology is a trending buzzword among various environments. Cloud can access data and related files from any location and from any device at any time with an internet connection. With the rapid development of cloud storage, enterprises tend to out-source their sensitive data to remote cloud service providers in pay-per-use.As the working process is changing to flexible and remote working, it is essential to provide user-related data access to users, even when they are not at a workplace. Improved data security is another asset of cloud computing. With call it stub-reserved attack. Furthermore, we propose a secure data deduplication scheme with efficient re-encryption based on the convergent all-or-nothing transform (CAONT) and randomly sampled bits from the Bloom filter. Due to the intrinsic property of one-way hash function, our scheme can resist the stub-reserved attack and guarantee the data privacy of data owners' sensitive data. Moreover, instead of re-encrypting the entire package, data owners are only required to re-encrypt a small part of it through the CAONT, thereby effectively reducing the computation overhead of the system. Finally, security analysis and experimental results show that our scheme is secure and efficient in re- encryption.
However, cloud service providers may store plentiful and repetitive data (such as movies, music and genome data), which inevitably incurs a mass of redundant storage and backup space, consequently to cost a vast amount of computing and management overhead during its whole life cycle.

## II. RELATED WORK

Numerous researchers have devoted considerable attention to the problem of how to support data deduplication under ciphertext. CE scheme is the first clever solution for data deduplication over encrypted data [17]. The main idea is as follows: a user encrypts and decrypts some sensitive data with a convergent key, which is derived by hashing these data. Since the convergent key and data are deterministic, the identical data is deterministically encrypted to the same ciphertext, no matter who encrypts them. Thus, this allows cloud service providers to perform deduplication over ciphertexts. However, the CE scheme is inherently vulnerable to the brute-force dictionary attacks. In order to solve this problem, Bellare et al. [35] proposed the DupLESS scheme, in which a user obtains the key from a dedicated keyserver via an oblivious PRF (OPRF) protocol. The OPRF mechanism is used to "blind" the fingerprint. Based on the RSA mechanism, the key server is configured with a system-wide public/private key pair. This enables the key

|| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | SJIF: 8.423|| Volume 12, Special Issue 2, April 2023 ||

8th National Conference on Advanced Computing Technologies (NCACT'23)

Organized by

Department of CSE, Velammal Engineering College, Chennnai, India

server to return the MLE key without knowing the original fingerprint. The rate-limits are used in the key generation requests and can be efficient against the brute-force attacks. If the key-server is secure, the encryption key appears to be derived from a random space. Shin et al. [39] extended the predicate encryption scheme in data deduplication. However, this scheme only supports the requirement of singleuser data deduplication. By introducing the additional tag checking mechanism, Bellare et al. proposed the randomized convergent encryption (RCE) scheme [38]. After decrypting the ciphertext, the user uses the plaintext to generate the tag and compare it with the corresponding tag. If tags are consistency, the user accepts the ciphertext; else, rejects it. Thus, the RCE scheme guarantees the integrity of users' data. However, these schemes suffer from security flaws with respect to user revocation. If the revoked users keep the convergent key, they can access the plaintext without permission. Thus, the confidentiality of users' sensitive data cannot be guaranteed.

## III. PROPOSED SYSTEM

We introduce the distributed cloud storage servers into deduplication systems to provide better fault tolerance. To further protect data confidentiality, the secret sharing technique is utilized, which is also compatible with the distributed storage systems. In more details, a file is first split and encoded into fragments by using the technique of secret sharing, instead of encryption mechanisms. These shares will be distributed across multiple independent storage servers. Furthermore, to support deduplication, a short cryptographic hash value of the content will also be computed and sent to each storage server as the fingerprint of the fragment stored at each server. Only the data owner who first uploads the data is required to compute and distribute such secret shares, while all following users who own the same data copy do not need to compute and store these shares any more.

## IV. EXISTING SYSTEM

A number of deduplication systems have been proposed based on various deduplication strategies such as client-side or server-side deduplications, file-level or block-level deduplications. Bellare et al formalized this primitive as message-locked encryption, and explored its application in space efficient secure outsourced storage. There are also several implementations of convergent implementations of different convergent encryption variants for secure deduplication. Li addressed the keymanagement issue in block-level deduplication by distributing these keys across multiple servers after encrypting the files. showed how to protect data confidentiality by transforming the predictable message into a unpredictable message. Data reliability is actually a very critical issue in a deduplication storage system because there is only one copy for each file stored in the server shared by all the owners. Most of the previous deduplication systems have only been considered in a single-server setting. The traditional deduplication methods cannot be directly extended and applied in distributed and multi-server systems.

## V. LITERATURE REVIEW

### 1.ENCRYPTED DEDUPLICATION STORAGE USING STORAGE

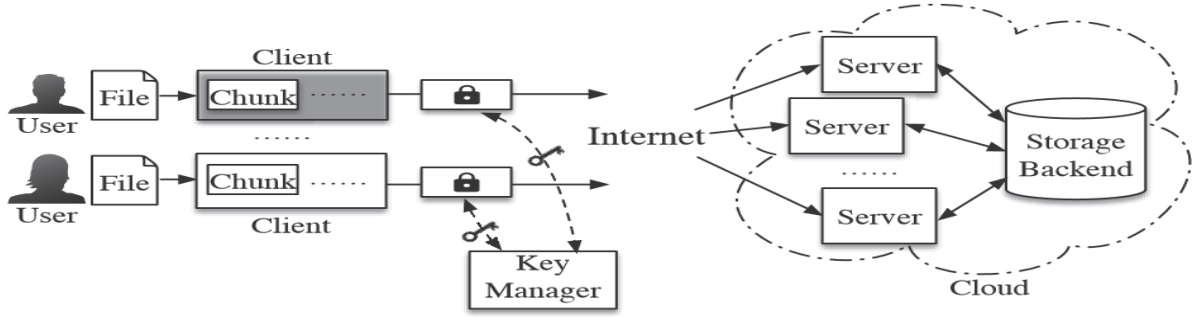
Rekeying refers to an operation of replacing an existing key with a new key for encryption. It renews security protection, so as to protect against key compromise and enable dynamic access control in cryptographic storage. However, it is non-trivial to realize efficient rekeying in encrypted deduplication storage systems, which use deterministic content-derived encryption keys to allow deduplication on ciphertexts. We design and implement REED, a rekeying-aware encrypted deduplication storage system. REED builds on a deterministic version of all-or-nothing transform (AONT), such that it enables secure and lightweight rekeying, while preserving the deduplication capability. We propose two REED encryption schemes that trade between performance and security, and extend REED for dynamic access control. We implement a REED prototype with various performance optimization techniques. Our tracedriven testbed evaluation shows that our REED prototype maintains high performance and storage efficiency.

|| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | SJIF: 8.423|| Volume 12, Special Issue 2, April 2023 ||

8th National Conference on Advanced Computing Technologies (NCACT'23)

Organized by

Department of CSE, Velammal Engineering College, Chennnai, India




## 2.IMPROVED REABILITY WITH SECURE DISTRIBUTED DUPLICATION SYSTEM

Deduplication system improves storage utilization while reducing reliability. Furthermore, the challenge of privacy for sensitive data also arises when they are outsourced by users to cloud. Aiming to address the above security challenges, this paper makes the first attempt to formalize the notion of distributed reliable deduplication system. We propose new distributed deduplication systems with higher reliability in which the data chunks are distributed across multiple cloud servers. The security requirements of data confidentiality and tag consistency are also achieved by introducing a deterministic secret sharing scheme in distributed storage systems, instead of using convergent encryption as in previous deduplication systems. Security analysis demonstrates that our deduplication systems are secure in terms of the definitions specified in the proposed security model. As a proof of concept, we implement the proposed systems and demonstrate that the incurred overhead is very limited in realistic environments.

### FIRST LEVEL OF DATA DUPLICATION:

- To support efficient duplicate check, tags for each file will be computed and are sent to S-CSPs.
- More precisely, the user firstly computes and sends the file tag $\phi F$ = TagGen(F) to S-CSPs for the file duplicate check.
- If a duplicate is found the user computesand sendsit to a server via a secure channel.
- Otherwise if no duplicate is found the process continues,i.e secret sharing scheme runs and the user will upload a file to CSP.
- To download a file the user will use the secret shares and download it from the SCSP's .
- This approach provides fault tolerance and allows the user to remain accessible even if any limited subsets of storage servers fail.

|| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | SJIF: 8.423|| Volume 12, Special Issue 2, April 2023 ||

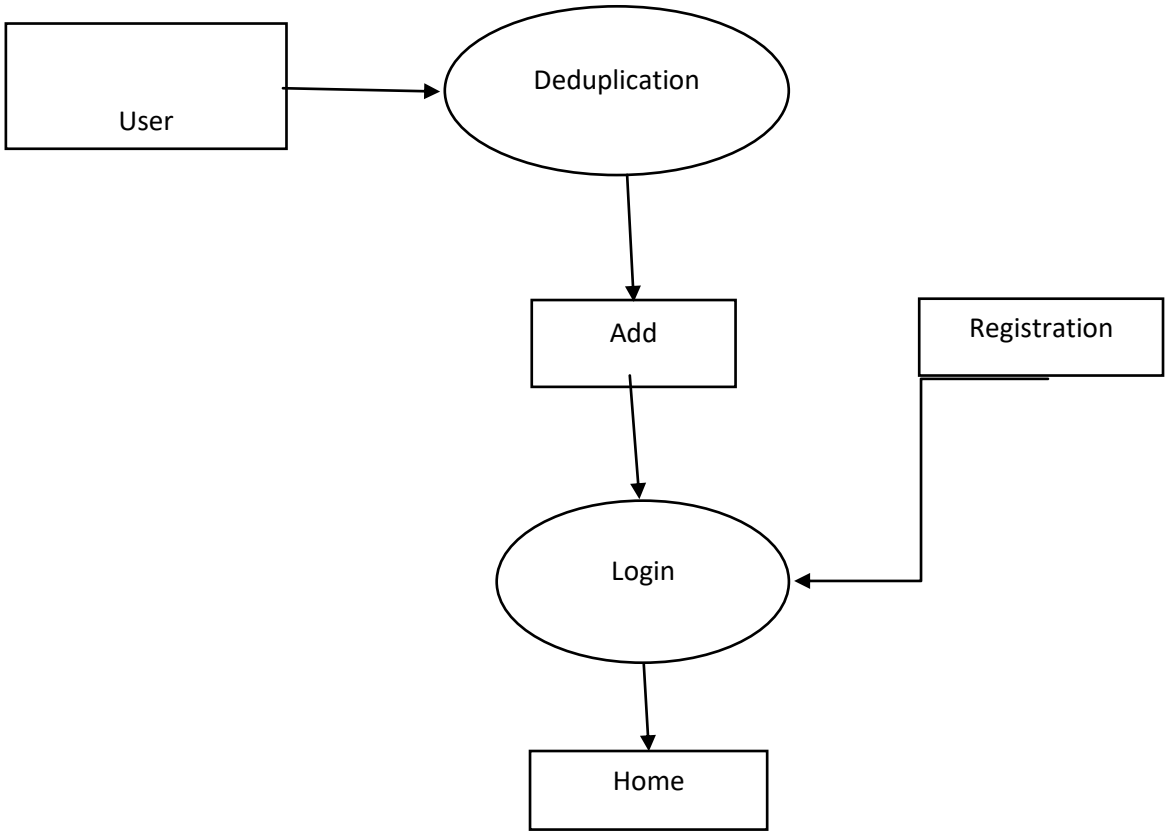
8th National Conference on Advanced Computing Technologies (NCACT'23)

Organized by

Department of CSE, Velammal Engineering College, Chennnai, India

**FLOW DIAGRAM**

LEVEL 0



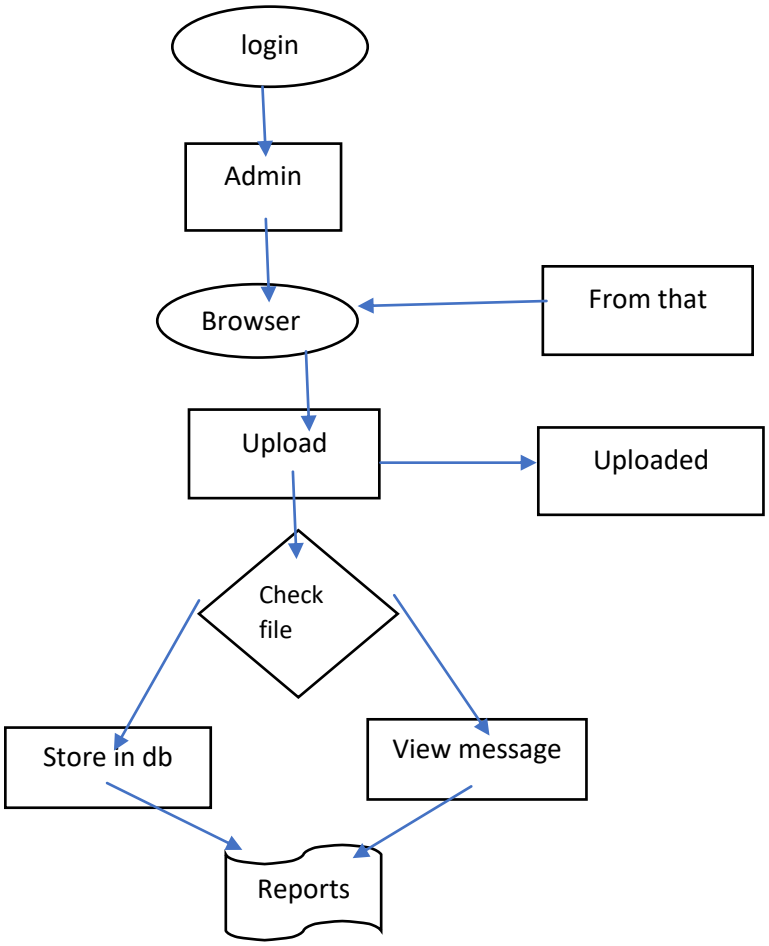
|| e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | SJIF: 8.423|| Volume 12, Special Issue 2, April 2023 ||

8th National Conference on Advanced Computing Technologies (NCACT'23)

Organized by

Department of CSE, Velammal Engineering College, Chennnai, India

Level 1




## VII. CONCLUSION

Managing encrypted data with deduplication is most significant in practice for running a cloud storage service which is secure and dependable, especially for data processes. We proposed the distributed deduplication systems to improve the reliability of data while achieving the confidentiality of the user outsourced data without an encryption mechanism. Four constructions were proposed to support file- level deduplication. The security of tag consistency and integrity were achieved.

## REFERENCES

1) M. Gerla, J. Weng, and G. Pau, "Pics-on-wheels: Photo surveillance in the vehicular cloud," International Conference on Computing, Networking and Communications, pp. 1123–1127, 2013.
2) X. Chen, J. Li, J. Ma, Q. Tang, and W. Lou, "New algorithms for secure outsourcing of modular exponentiations," IEEE Trans. Parallel Distrib. Syst., vol. 25, no. 9, pp. 2386–2396, 2014.
3) H. Yuan, X. Chen, T. Jiang, X. Zhang, Z. Yan, and Y. Xiang, " Dedupdum: Secure and scalable data deduplication with dynamic user management," Inf. Sci., vol. 456, pp. 159–173, 2018.
4) H. Huang, X. Chen, Q. Wu, X. Huang, and J. Shen, "Bitcoin based fair payments for outsourcing computations of fog devices," Future Generation Comp. Syst., vol. 78, pp. 850–858, 2018.
5) IDC. (2014) The digital universe of opportunities: Rich data and the increasing value of the internet of things. [Online]. Available: https://www.emc.com/leadership/ digital universe/2014iview/index.html
6) W. J. Bolosky, S. Corbin, D. Goebel, and J. R. Douceur, "Single instance storage in windows 2000," in Conference on Usenix Windows Systems Symposium, 2000.

INNO SPACE
SJIF Scientific Journal Impact Factor

SJIF: 8.423

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INTERNATIONAL CENTRE

निस्केयर NISCAIR

# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

## IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details