IJARETY

International Journal of Advanced Research in Education and TechnologY (IJARETY)

Volume 12, Issue 3, May-June 2025

Impact Factor: 8.152

www.ijarety.in editor.ijarety@gmail.com

International Journal of Advanced Research in Education and TechnologY(IJARETY)

ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 3, May-June 2025 ||

DOI:10.15680/IJARETY.2025.1203075

Efficient Secure Data Deduplication with Popularity Sensitive Tag Randomization

Mr. A. Rajesh [1], M Ganesh[2], J Sravan Kumar[3], M Navateja[4]

Associate Professor, Department of CSE, Guru Nanak Institute of Technology, Hyderabad, Telangana, India[1]

Student, Department of CSE, Guru Nanak Institute of Technology, Hyderabad, Telangana, India[2, 3, 4]

ABSTRACT: It is non-trivial to provide semantic security for user data while achieving deduplication in cloud storage. Some studies deploy a trusted party to store deterministic tags for recording data popularity, then provide different levels of security for data according to popularity. However, deterministic tags are vulnerable to offline brute-force attacks. In this project, we first propose a popularity-based secure deduplication scheme with fully random tags, which avoids the storage of deterministic tags

KEYWORDS: Data Deduplication, Secure Storage, Tag Randomization, Data Privacy, Cloud Security

I. INTRODUCTION

The rapid development of cloud computing, more and more individuals or enterprises choose to outsource data to the cloud server. The storage of large volumes of data brings huge overheads to the cloud service providers. Deduplication is an effective way to detect and eliminate redundant copies over clouds. The cloud server only stores the unique data after deduplication to reclaim a lot of storage space. Due to the insecure network environment, users tend to outsource encrypted data to prevent data privacy from being snooped on. However, conventional encryption algorithms aim to provide semantic security for user data. In other words, the encrypted data are indistinguishable from random bits. This property hinders data deduplication since the same messages will be encrypted into indistinguishable ciphertexts. Convergent encryption (CE) is the first attempt to achieve encrypted deduplication. The encryption keys in CE are derived from the data content, so it is a deterministic encryption algorithm and can make sure that identical messages could be encrypted into identical ciphertexts. Nevertheless, CE provides confidentiality only for unpredictable data (the message space cannot be exhausted) . For predictable data (the message space can be exhausted), CE is vulnerable to offline brute-force attacks.

II. LITERATURE SURVEY

Title: Enhanced secure thresholded data deduplication scheme for cloud storage
Year: 2018
Author: J. Stanek and L. Kencl.
Description: As more corporate and private users outsource their data to cloud storage, recent data breach incidents make end-to-end encryption increasingly desirable. Unfortunately, semantically secure encryption renders various cost-effective storage optimization techniques, such as data deduplication, ineffective. On this ground Stanek et al. [1] introduced the concept of “data popularity” arguing that data known/owned by many users do not require as strong protection as unpopular data; based on this, Stanek et al. presented an encryption scheme, where the initially semantically secure ciphertext of a file is transparently downgraded to a convergent ciphertext that allows for deduplication as soon as the file becomes popular. In this paper we propose an enhanced version of the original scheme. Focusing on practicality, we modify the original scheme to improve its efficiency and emphasize clear functionality. We analyze the efficiency based on popularity properties of real datasets and provide a detailed performance evaluation, including comparison to alternative schemes in real-like settings. Importantly, the new scheme moves the handling of sensitive decryption shares and popularity state information out of the cloud storage, allowing for improved security notion, simpler security proofs and easier adoption. We show that the new scheme is secure under the Symmetric External Diffie-Hellman assumption in the random oracle model.

Title: A secure deduplication scheme based on data popularity with fully random tags
Year: 2021
Author: G. Ha, H. Chen, C. Jia, R. Li and Q. Jia.

SSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 3, May-June 2025 ||

DOI:10.15680/IJARETY.2025.1203075

Description: It is difficult to provide semantic security for user data while using deduplication to save storage space in cloud storage. Some studies attempt to provide different levels of security for data according to their popularity for a reasonable trade-off between security and efficiency. However, existing schemes generally need a trusted third party to store deterministic data tags to record data popularity. If the trusted third party is compromised by adversaries, the deterministic tags will expose data information. In this paper, we propose a popularity-based secure deduplication scheme with fully random tags, which does not need to store deterministic tags. Our solution is using the homomorphic encryption to generate comparable random tags to record data popularity and using binary search to reduce time complexity of tag comparison to logarithmic time. Besides, we also design a proof of ownership protocol based on homomorphic encryption to prevent adversaries with only the data tag from tampering with the data popularity, which are not considered in the existing popularity-based schemes. We implement our scheme for system efficiency evaluation. Compared with the scheme of Stanek et al., our scheme has a slight improvement in encryption efficiency.

Title: Secure deduplication of encrypted data: Refined model and new constructions
Year: 2018
Author: J. Liu, L. Duan, Y. Li and N. Asokan.
Description: Cloud providers tend to save storage via cross-user deduplication, while users who care about privacy tend to encrypt their files on client-side. Secure deduplication of encrypted data (SDoE) which aims to reconcile this apparent contradiction is an active research topic. In this protocols and prove their security in our model. We evaluate their deduplication effectiveness via simulations with realistic datasets.

Cloud storage services are very popular. Providers of cloud storage services routinely use cross-user deduplication to save costs: if two or more users upload the same file, the storage provider stores only a single copy of the file. Users concerned about privacy of their data may prefer encrypting their files on clientside before uploading them to cloud storage. This thwarts deduplication since identical files are uploaded as completely different ciphertexts. Reconciling deduplication and encryption has been a very active research topic . One proposed solution is convergent encryption (CE) which derives the file encryption key solely and deterministically from the file contents. As a result, identical files will always produce identical ciphertexts given identical public parameters. Unfortunately, a server compromised by the adversary can perform an offline brute-force guessing attack over the ciphertexts, due to the deterministic property of CE.paper, we propose a formal security model for SDoE. We also propose two single-server SDoE

Existing System

- Existing popularity-based encrypted deduplication schemes need to deploy a trusted third party to store deterministic tags for recording data popularity.
- However, if the trusted party is compromised by adversaries, the deterministic tags will be vulnerable to offline brute-force attacks, which is a serious security vulnerability.
- Besides, we find the “popularity tamper attack” in existing popularity-based schemes. Since unpopular data are usually encrypted randomly in popularity-based schemes, it is difficult for the server to verify the data ownership for users.
- In other words, it is difficult to design proof-of-ownership (PoW) protocols for unpopular data in popularity-based schemes.

Existing System Disadvantages

- Not achieve efficient data encryption and key update.
- Less storage efficiency.
- Semantic security and data deduplication seem to be irreconcilable

Proposed System

- To achieve scalability and updatability, we introduce the multi-key homomorphic proxy re-encryption (MKH-PRE) to design a multi-tenant scheme.
- Users in different tenants generate tags using different key pairs, and the cross-tenant tags can be compared for equality. Meanwhile, our multi-tenant scheme supports efficient key updates.
- We give comprehensive security analysis and conduct performance evaluations based on both synthetic and real-world datasets.
- The results show that our schemes achieve efficient data encryption and key update, and have high storage efficiency.

International Journal of Advanced Research in Education and TechnologY(IJARETY)

SSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 3, May-June 2025 ||

DOI:10.15680/IJARETY.2025.1203075

Proposed System Advantages

- Our schemes achieve efficient data encryption and key update.
- High storage efficiency.
- The storage efficiency of the popularity-based deduplication scheme.

III. SYSTEM ARCHITECTURE

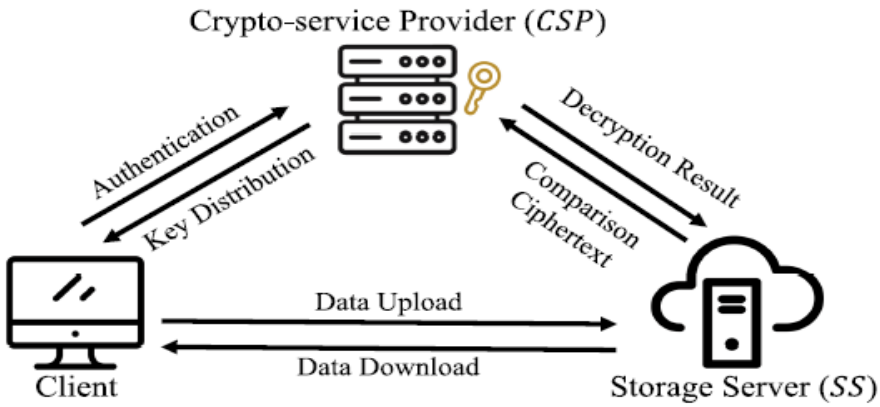

Fig System Architecture

It is non-trivial to directly perform equality-testing on random tags. We use the homomorphic ciphertexts of deterministic tags as random tags to resolve this issue. HE supports equivalence comparison on ciphertexts without decryption. So, we can perform equality-testing on random tags. Moreover, HE can also be used to resist offline brute-force attacks, in that it is probabilistic rather than deterministic. As shown in Fig. 3, our scheme deploys a crypto-service provider (CSP) to manage the key pair of HE and perform the homomorphic decryption. When the storage server (SS) performs the tag equality-testing, it sends the homomorphism subtractions between random tags to CSP. The latter decrypts them and returns the comparison results to SS. The second challenge is how to improve the efficiency of the equality-testing of random tags. The linear time complexity of the tag comparison is unacceptable when there are a large number of random tags. Our solution is to use binary search. Due to the property of HE, SS can get the algebraic relationship between any two random tags by interacting with CSP

IV. METHODOLOGY

Modules Name:

- User Interface Design
- Client
- Storage Server (SS) or Cloud Server
- Crypto-Service Provider (CSP)
- TPA

1. User Interface Design

To connect with server user must give their username and password then only they can able to connect the server. If the user already exits directly can login into the server else, user must register their details such as username, password, Email id, City and Country into the server. Database will create the account for the entire user to maintain upload and download rate. Name will be set as user id. Logging in is usually used to enter a specific page. It will search the query and display the query.

2. Client

The client outsources user data to SS to save local storage overhead. To maintain privacy, the client outsources encrypted data to SS. Our method does not need the involvement of clients and supports dynamic insertion and

SSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 3, May-June 2025 ||

DOI:10.15680/IJARETY.2025.1203075

deletion of tags. Besides, we find the popularity tamper attack in the popularity-based deduplication and design a PoW protocol based on HE to resist it. Only the users with the whole data content could increase the count of data owners

3. Storage Server or Cloud Server

The storage server provides data storage services for multiple users and performs cross-user deduplication to save storage space. To this end, we design a PoW protocol based on HE to resist the popularity tamper attack without leaking any data information. The idea of our PoW is to let clients compute the hashes of some randomly sampled challenge blocks and then encrypt these hashes with HE. These encrypted hashes are used as the proofs for data ownership. SS can verify whether the proofs are valid through the interaction with CSP. Since the proofs are randomly encrypted, they do not reveal any data information.

4. Crypto-Service Provider:

The crypto-service provider is independent of clients and the storage server. It is responsible for managing the homomorphic encryption key pair. It authenticates clients and distributes the homomorphic encryption public key to SS and all authenticated clients. our scheme deploys a crypto-service provider to manage the key pair of homomorphic encryption and perform the homomorphic decryption.

Note that crypto-service provider can be implemented by the third-party external cryptographic service, where a cryptographic server performs some cryptographic operations for applications. Many cryptography-based schemes and well-resourced enterprises deploy external cryptographic services.

5. TPA

This is the Third module in our project where TPA working process. TPA has to register then login with valid username and password. After login successful he can do some operations such as new files request, update files requests, new audit files, new verified files form cloud server and updated file form cloud servers. Secure cloud storage protocols are publicly verifiable if an audit can be performed by any Third Party Auditor (TPA) using public parameters; or privately verifiable if an auditor needs some secret information of the client. The entities involved in a secure cloud storage protocol and the interaction among them are working. Audit the proof provided by the CSP and inform the DO of the result. In this scheme, the smart contracts deployed on the blockchain and consensus nodes cooperate to perform audit tasks.

V. EXPERIMENTAL RESULTS

Home page:

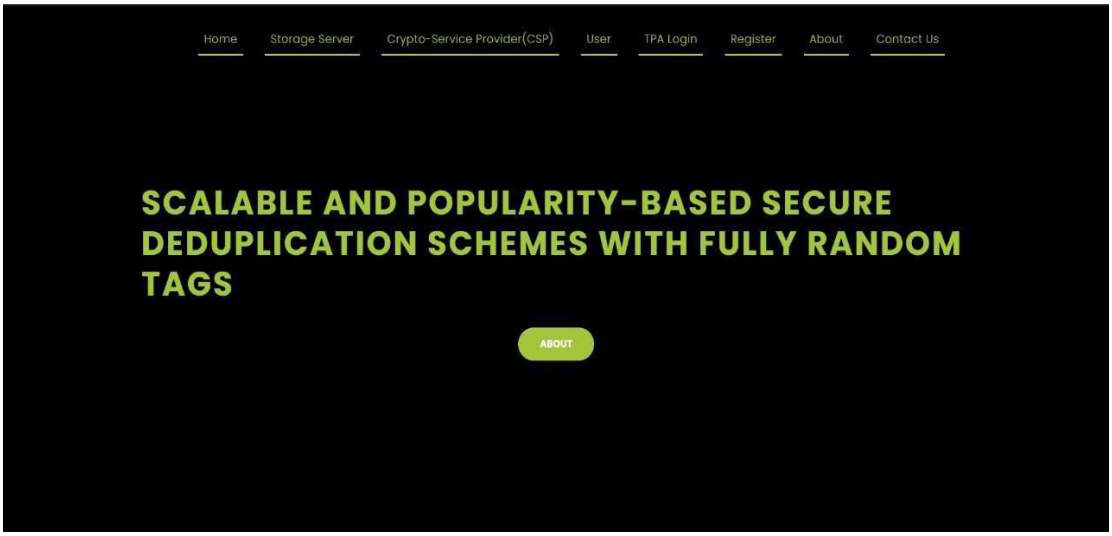

Fig: 2 Home Page

International Journal of Advanced Research in Education and TechnologY(IJARETY)

SSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 3, May-June 2025 ||

DOI:10.15680/IJARETY.2025.1203075

Registration Page:

Fig 3 Registration Page

The registration page is a user interface designed to facilitate the process of creating a new account. It typically features input fields for gathering essential information from the user, such as their name, email address, and a secure password.

Login Page:

Fig: 4 login Page

Fig: 7 Result Page

SSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 3, May-June 2025 ||

DOI:10.15680/IJARETY.2025.1203075

V. CONCLUSION

In this project, we first propose a single-tenant popularity-based encrypted deduplication scheme with fully random tags. We use HE to generate random tags, avoiding storing deterministic tags to record data popularity. Besides, we reduce the time complexity of tag equality-testing by the binary search in the AVL tree. We also design a PoW protocol to resist the popularity tamper attack. For scalability and key rotation, we expand our single-tenant scheme to a multi-tenant scheme by introducing MKH-PRE. In the multitenant scheme, users in different tenants use different HE key pairs to generate data tags, while the server could record the cross-tenant data popularity. The multi-tenant scheme also supports key rotation based on the proxy re-encryption of MKH-PRE. We implement prototypes of our schemes and evaluate their performances. The results show that our schemes have high storage efficiency and achieve efficient data encryption and key update.

VI. FUTURE ENHANCEMENT

We implement prototypes of our schemes and evaluate their performances. The results show that our schemes have high storage efficiency and achieve efficient data encryption and key update. The time overheads include the overheads of random tag generation, data encryption, and data communication. The overheads of the ciphertext validation include the overheads of generating the deterministic tag, once HE encryption, and once HE decryption, which are close to the overheads of the data

REFERENCES

1. J. R. Douceur, A. Adya, W. J. Bolosky, P. Simon and M. Theimer, "Reclaiming space from duplicate files in a serverless distributed file system", Proc. 22nd Int. Conf. Distrib. Comput. Syst., pp. 617-624, 2002.
2. M. Bellare, S. Keelveedhi and T. Ristenpart, "DupLESS: Server-aided encryption for deduplicated storage", Proc. Usenix Conf. Secur., pp. 179-194, 2013.
3. J. Stanek and L. Kencl, "Enhanced secure thresholded data deduplication scheme for cloud storage", IEEE Trans. Dependable Secure Comput., vol. 15, no. 4, pp. 694-707, Jul./Aug. 2018.
4. P. Puzio, R. Molva, M. Önen and S. Loureiro, "PerfectDedup: Secure data deduplication", Proc. Int. Workshop Data Privacy Manage., pp. 150-166, 2015.
5. S. Halevi, D. Harnik, B. Pinkas and A. Shulman-Peleg, "Proofs of ownership in remote storage systems", Proc. ACM Conf. Comput. Commun. Secur., 2011.
6. J. Xu, E. C. Chang and J. Zhou, "Weak leakage-resilient client-side deduplication of encrypted data in cloud storage", Proc. 8th ACM SIGSAC Symp. Inf. Comput. Commun. Secur., pp. 195-206, 2013.
7. T. Jiang, X. Chen, Q. Wu, J. Ma, W. Susilo and W. Lou, "Secure and efficient cloud data deduplication with randomized tag", IEEE Trans. Inf. Forensics Secur., vol. 12, no. 3, pp. 532-543, Mar. 2017.
8. M. Abadi, D. Boneh, I. Mironov, A. Raghunathan and G. Segev, "Message-locked encryption for lockdependent messages", Proc. 33rd Annu. Cryptol. Conf., pp. 374-391, 2013.
9. G. Ha, H. Chen, C. Jia, R. Li and Q. Jia, "A secure deduplication scheme based on data popularity with fully random tags", Proc. IEEE 20th Int. Conf. Trust Secur. Privacy Comput. Commun., pp. 207-214, 2021. 10. M. Bellare and S. Keelveedhi, Message-Locked Encryption and Secure Deduplication, Berlin, Germany: Springer, 2013.
10. J. Liu, N. Asokan and B. Pinkas, "Secure deduplication of encrypted data without additional independent servers", Proc. 22nd ACM SIGSAC Conf. Comput. Commun. Secur., pp. 874-885, 2015.

International Journal of Advanced Research in Education and Technology

ISSN: 2394-2975 Impact Factor: 8.152

www.ijarety.in editor.ijarety@gmail.com