Journal of Science and Technology
ISSN: 2456-5660 Volume 8, Issue 07 (July -2023)
www.jst.org.in *DOI:https://doi.org/10.46243/jst.2023.v8.i07.pp152- 157*

An Efficient Single Instance Scheme With User Authentication To Cloud Data

Boini Vandana[1] |P.Sathish[2] | Dr.V.Bapuji[3]

[1]Department of MCA, Vaageswari College of Engineering, Karimnagar.
[2] Assistant Professor, Department of MCA, Vaageswari College of Engineering, Karimnagar.
[3] Professor & HoD, Department of MCA, Vaageswari College of Engineering, Karimnagar.

To Cite this Article

Boini Vandana |P.Sathish | Dr.V.Bapuji , "**An Efficient Single Instance Scheme With User Authentication To Cloud Data** " *Journal of Science and Technology, Vol. 08, Issue 07,- July 2023, pp152-157*

Article Info

Received: 06-06-2023 Revised: 07-07-2023 Accepted: 16-07-2023 Published: 27-07-2023

ABSTRACT

Cloud Storage is a computer data storage method where digital data is stored on servers in off-site locations, managed by a third-party provider. This enables organizations to store, access, and maintain data without owning and operating their own data centers. Cloud storage is scalable, allowing organizations to expand or reduce their data footprint depending on their needs. Users upload data to servers via internet, which is saved on a virtual machine on a physical server. Cloud providers often spread data to multiple virtual machines in global data centers to maintain availability and redundancy. Google Cloud offers various scalable options for organizations to store their data in the cloud. The widespread use of cloud computing has made data sharing and storage more accessible, but concerns about data integrity, efficiency, and privacy remain. Duplication, a popular method of data compression, is used to reduce duplicate copies of data in cloud storage.

However, data duplication also raises security and privacy concerns, as users' confidential data is vulnerable to attacks from insiders and outsiders. Traditional solutions for duplication, based on convergent encryption, provide confidentiality but do not maintain duplicate checks based on differential permissions. This paper proposes an approved data duplication plan that counts the number of users with differential privileges in the duplicate check.Users with differential privileges are added to the duplicate check, and files are encrypted with differential privilege keys to maintain stronger security. Users can only access files marked with matching privileges for copy checks.

A third-party auditor can confirm file occurrence after duplication in the cloud, ensuring timely uploads. This paper offers advantages for both storage providers and users through duplication systems and auditing methods.

KEYWORDS: *Cloud storage, Dropbox, Mozy, Perfect Hashing, Storage, Encryption, Attacks, Privacy*

1.INTRODUCTION

There will be 4 billion digital files stored in the cloud by 2020. Costs associated with management, upkeep, and handling are substantial. By keeping redundant information only once, information duplication techniques try to get rid of duplicate data. However, outsourcing sensitive data necessitates its encryption, which makes duplication attempts more difficult. Numerous solutions have been put out to deal with this problem, however they are hindered by things like brute-force attacks and the storage capacity limitations of the cloud.

1950 saw the introduction of lossy and lossless data reduction techniques. Techniques for data duplication were created in the early 2000' s after 1990 saw the introduction of compression methods like delta compression. These techniques aid in reducing data repetition and deal with multimedia capacity constraints.

By 2025, the size of the world's data sphere is predicted to be 185 zettabytes, necessitating the need of massive data storage. Although duplication technology might save server and user costs, improperly verified permissions can lead to security issues. In order to enhance user permission checks and prevent security issues in SADS, a secure encrypted file duplication technique with permission is needed.
Cloud storage can duplicate user-encrypted files using the EFDSP scheme without erasing the owner's permission.

A service platform architecture called cloud computing provides on-demand access to resources in the form of X as a Service. The server less and effective encrypted duplication (SEED) for mobile cloud storage services, as well as the secure authorized duplication schemes for hybrid clouds (SADS). It makes use of secure duplication at the source, cipher text and tag randomization, and bi linear pairing-based encryption.

A service platform architecture called cloud computing provides on-demand access to resources in the form of X as a Service. The server less and effective encrypted duplication (SEED) for mobile cloud storage services, as well as the secure authorized duplication schemes for hybrid clouds (SADS). It makes use of safe block level duplication, cipher text and tag randomization, and bi linear pairing-based encryption. Provable security is provided through SEED, which allows for cross-user duplication of encrypted data [9]. The method ensures secure duplication at the block level by effectively identifying popular data segments using a Perfect Hash Function. providing block-level secure duplication.

The purposes of data duplication It removes items (copies) that already exist in the data collection and compares objects (often files or blocks).

Blocks that are not unique are removed during the duplication process.

1. *Separate the input data into sections or "chunks."*
2. *Determine a hash value for every data block.*
3. *Use these values to ascertain whether the same block of data has previously been stored in another block.*
4. *Substitute a reference to the database object for the duplicate data.*

The duplicates can be located and removed after the data has been chunked, allowing for the creation of an index. Data is only stored once per instance.Once the data is chunked, an index can be created from the results, and the duplicates can be found and eliminated. Only single instance of data is stored. Data duplication is important because it lowers your need for storage space, helps you save money, and uses less bandwidth to transfer data to and from remote storage locations. In some cases, data duplication can reduce the amount of storage needed by up to 95%, albeit the sort of data you are trying to duplicate may have an impact on your specific duplication ratio.

2 EXISTING SYSTEM

Various schemes have been proposed to address this issue, but they face challenges like brute-force attacks and limited cloud supported storage capabilities. Duplication technology aims to reduce redundant data in cloud servers, saving both the server and user costs. Data deduplication techniques aim to prevent brute-force attacks and ensure data effectiveness[6].

Encrypted file duplication schemes can save storage space and network bandwidth,but security concerns arise when permissions are not properly checked. By storing only one replica for numerous identical pieces of data, deduplication technology helps cloud servers and users[7].
One copy of each submitted file is stored by cloud storage companies like Dropbox, Google Drive, and Mozy in order to conserve space. For reasons of privacy, corporate policy, and legal requirements, customers want data encryption.[2]. Existing industrial solutions fail in encrypted data duplication. Clients can validate server possession without requesting original data thanks to the newly introduced PDP paradigm.The model generates probabilistic proofs of possession by sampling random sets of blocks from the server reducing I/O costs.

The challenge/response protocol transmits a tiny, consistent quantity of data, minimizing network communication, and the client keeps a constant amount of metadata to validate the proof. It is demonstrated that the two PDP systems that are provided are more effective than earlier ones. Building systems that are effective and secure while enabling client data extraction from verification tests is a problem. The first approach, which is secure in the random oracle model and built from BLS signatures, has the quickest query and response times.

DISADVANTAGES OF EXISTING SYSTEM

1. Information integrity is compromised when hashes are used for querying information, as they can cause collisions and loss of integrity.
2. Security and privacy strategies should be carefully designed to prevent security breaches and personal information damage.

3. Fixed-size storage frameworks can cause capacity execution issues, requiring additional resources like memory, CPU, and data transfer capacity.

4. Additionally, backup machines may require different equipment to move and process information, potentially causing additional storage execution issues.
5. Deterministic guarantees lack data ownership, cannot be given without access to all blocks, and only provide sampling between blocks.
6. Security is questionable, and source authentication techniques do not apply to provable data possession.

3 PROPOSED SYSTEM

Duplication is an emerging technology that can be developed using efficient algorithms. The difference in a single character in a string has minimal impact on data, unlike digits. The cloud's analysis power needs to be advanced to identify problems and solutions. Optimized algorithms are crucial for an expanding global cloud data set. One area to focus on is developing a duplication procedure for multilingual datasets, allowing for comparisons to detect different scripts based on the same data.

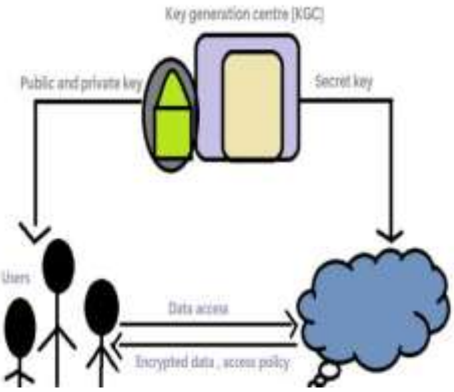

FIG : 1 System Architecture

The project proposes a cloud platform that allows users or new registrations to access its facilities. The platform will provide standard features such as uploading new files or reading or downloading already uploaded files. The platform will also implement duplication checks to verify if similar files or data are already present in the cloud storage.

If negative, the file will be uploaded in ciphered form with proof of ownership. The approval of ownership from the cloud administration will generate a convergent key based on the data provided.This key is generated using hash values based on data, and similar data will lead to a similar key.

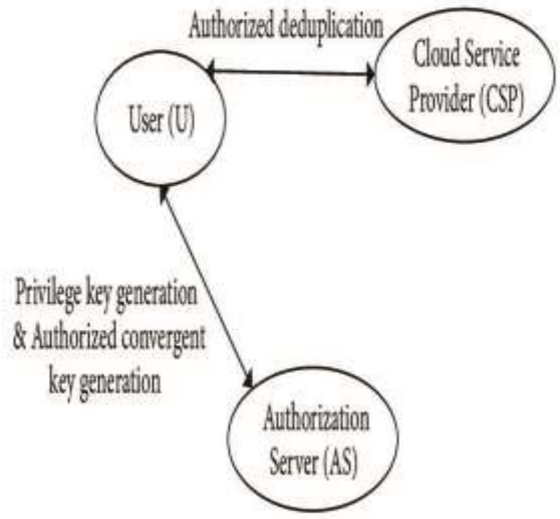
Authorized deduplication
User (U)
Cloud Service Provider (CSP)
Privilege key generation & Authorized convergent key generation
Authorization Server (AS)

FIG : 2 Process

To implement differential authorization, a token is generated with two inputs: the convergent key and the privileges associated with the user who uploaded the file. This token is provided to the user only upon the approval of proof of ownership. In a different scenario, if a duplicate file is uploaded, the user must provide proof of ownership. If verified, the second user will also receive a token to access their data based on their privileges.

This system differentiates between access to data across the cloud, maintaining privacy and differential access. The convergent key, token system, and data storage details are managed by an efficient database management system.

ADVANTAGES OF PROPOSED SYSTEM

1. *Our proposal offers storage, size reduction, efficient symmetric encryption, block-level duplication, no coordination, no storage overhead for unpopular data blocks.*
2. *Balances storage management and data security is crucial for organizations.*
3. *Improves data privacy and security by allowing users to specify who can access sensitive information, ensuring unauthorized access and compliance with privacy regulations.*
4. *Efficient data storage reduces footprint while maintaining information asset integrity and confidentiality.*
5. *Optimizing storage space, data privacy, security, governance, and regulatory compliance enhances organizational efficiency.*

4 RESULT AND ANALYSIS

As a result of the journal ,the deduplication with authorized user control in cloud for the data stored in the cloud storage is shown below in fig3.The plan enables CSP to control access permissions without sacrificing data security. The Bloom filter is utilized for effective duplicate checking. Data confidentiality, access control, tag consistency, and resistance to brute-force attacks are all achieved, according to security studies. Deduplication at the file and chunk levels is effective, according to performance tests, which lowers the cost of computation, communication, and storage.

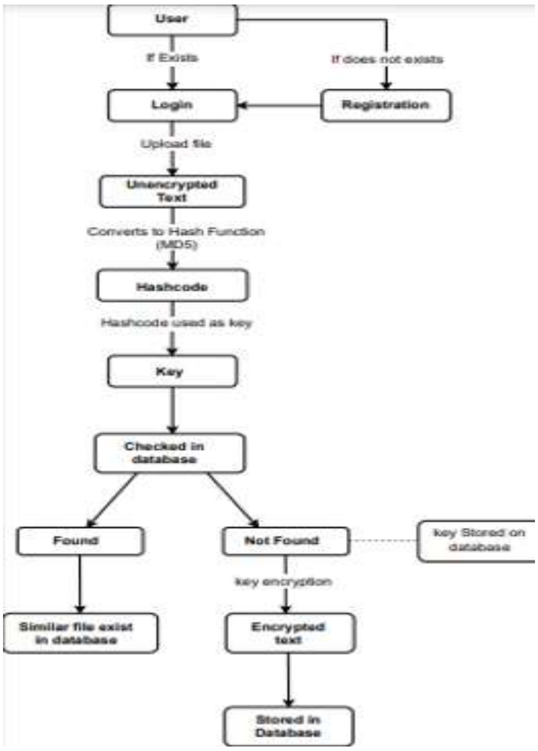

FIG:3 Resultant Deduplication

5 CONCLUSION

Encrypted data management with duplication is crucial for secure cloud storage services, particularly for data processes. future work should focus on efficient data ownership verification, scheme optimization, hardware acceleration for IOT devices, and flexible solutions for duplication and data access controlled by the owner or representative agent.

This paper proposal compresses data by removing duplicate copies, a technique widely used in cloud storage to save bandwidth and minimize storage space. convergent encryption is used to secure sensitive data during duplication, and the paper discusses data duplication authorization for better protection.

5 References

[1]. J. Stanek, A. Sorniotti, E. Androulaki, and L. Kencl, “A secure data duplication scheme for cloud storage,” in Financial Cryptography and Data Security - 18th International Conference, FC 2014, Christ Church, Barbados, March 3-7, 2014, Revised Selected Papers, 2014, pp. 99–118.
[2]. S. Keelveedhi, M. Bellare, and T. Ristenpart, “Dupless: Server-aided encryption for deduplicated storage,” in Proceedings of the 22th USENIX Security Symposium, Washington, DC, USA, August 14-16, 2013, 2013, pp. 179–194
[3]. M. Bellare, S. Keelveedhi, and T. Ristenpart, “Message-locked encryption and secure duplication,” in Advances in Cryptology - EUROCRYPT 2013, 32nd Annual International Conference on the Theory and Applications of Cryptographic Techniques, Athens, Greece, May 26-30, 2013.

[4]. Proceedings, 2013, pp. 296–312.X. Yang, R. Lu, K.-K. R. Choo, F. Yin, and X. Tang, “Achieving efficient and privacy-preserving cross-domain big data duplication in cloud,” IEEE Trans. Big Data, vol. PP, no. 99, pp. 1–1, 2017.

[5]. M. W. Storer, K. M. Greenan, D. D. E. Long, and E. L. Miller, “Secure data duplication,” in Proceedings of the 2008 ACM Workshop on Storage Security and Survivability, StorageSS 2008,Alexandria, VA, USA, October 31, 2008, 2008, pp. 1–10.

[6]. X. Yang, R. Lu, K. K. R. Choo, F. Yin and X. Tang, "Achieving Efficient and Privacy-Preserving Cross-Domain Big Data Deduplication in Cloud," in IEEE Transactions on Big Data, vol. 8, no. 1, pp. 73-84, 1 Feb. 2022, doi: 10.1109/TBDATA.2017.2721444.

[7]. Wang, Z., Gao, W., Yang, M. et al. Enabling Secure Data sharing with data deduplication and sensitive information hiding in cloud-assisted Electronic Medical Systems. Cluster Comput (2022)https://doi.org/10.1007/s10586-022- 03785-yLiang, X., Yan, Z., Chen, X., Yang, L. T., Lou, W., & Hou, T. (2019). Game Theoretical Analysis on Encrypted Cloud Data Deduplication. IEEE Transactions on Industrial Informatics, 15(10), 5778- 5789. https://doi.org/10.1109/TII.2019.2920402

[8]. Liang, X., Yan, Z., Chen, X., Yang, L. T., Lou, W., & Hou, T. (2019). Game Theoretical Analysis on Encrypted Cloud Data Deduplication. IEEE Transactions on Industrial Informatics, 15(10), 5778- 5789. https://doi.org/10.1109/TII.2019.2920402

[9]. Youngjoo Shin, Junbeom Hur, Dongyoung Koo, Joobeom Yun, "Toward Serverless and Efficient Encrypted Deduplication in Mobile Cloud Computing Environments", Security and Communication Networks, vol. 2020, Article ID 3046595, 15 pages, 2020. https://doi.org/10.1155/2020/3046595

[10]. DING, Wenxiu; YAN, Zheng; and DENG, Robert H.. Secure encrypted data deduplication with ownership proof and user revocation. (2017). Algorithms and architectures for parallel processing: 17th International Conference ICA3PP 2017, Helsinki, Finland, August 21-23, Proceedings. 297-312. Research Collection School Of Information Systems. Available at: https://ink.library.smu.edu.sg/sis_research/3788.