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.pp177- 183*

Secured And Efficient Data Duplication With Re-Encryption Techniques

Amirishetti Sukanya[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
Amirishetti Sukanya | P.Sathish | Dr.V.Bapuji, "Secured And Efficient Data Duplication With Re-Encryption Techniques " *Journal of Science and Technology, Vol. 08, Issue 07,- July 2023, pp177-183*
Article Info
Received: 04-06-2023 Revised: 05-07-2023 Accepted: 15-07-2023 Published: 27-07-2023

ABSTRACT

To get rid of duplicate copies and conserve bandwidth, data duplication is a vital approach in cloud storage. A convergent encryption strategy is suggested to encrypt data before outsourcing to maintain sensitive data confidentially. This work considers varied user privileges in duplicate checks to address authorized data duplication. In a hybrid cloud architecture, it introduces novel duplication structures that facilitate authorized duplicate checking. The proposed authorized duplicate check technique is constructed and tested, and security analysis proves the scheme's security. When compared to standard operations, the proposed system has a low overhead. Before outsourcing data into the cloud, encryption measures are frequently employed to ensure secrecy, however, commercial storage providers are hesitant to utilize encryption due to the possibility of different cipher texts. Convergent encryption, which protects data secrecy while permitting duplication, has been offered as a solution to these problems. Systems for cloud storage have difficulty maintaining duplication, dependability, and confidentiality. Duplication is a method for maximizing the use of storage resources by preventing the creation of duplicate copies of the same material. Both single-cloud duplication architecture and multi-cloud duplication architecture are categories for it.

KEYWORDS: Encryption, Storage, Authorized, Architecture, Duplication, Secrecy, Security.

I INTRODUCTION

Cloud technology allows for data access from any location with an internet connection, it improves data security and collaboration. It is essential to offer user-related data access as flexible and remote working grow increasingly common[2]. Based on the convergent all-or-nothing transform (CAONT) and randomly selected bits from the

Bloom filter, a safe data duplication strategy is suggested. This system can withstand stub-reserved assaults and guarantee data owners' privacy. Data owners just need to re-encrypt a tiny portion of the package instead of the complete package, which reduces the computational burden on the system.

The scheme's security and effectiveness in re-encryption are shown through security analysis and experimental findings. Compared to conventional data storage systems, cloud technology offers improved data safety via antivirus software, encryption techniques, power outages, and human mistake. The lightweight re-keying-aware encrypted duplication technique (REED), which is the subject of the paper's attention, is found to be susceptible to stub-reserved attacks[9].

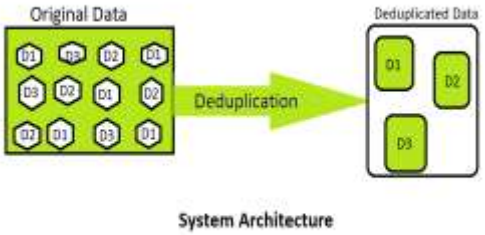

Fig.1.Deduplication reduces the amount of stored data

In general, cloud technology benefits from improved data storage, security, and collaboration. Before outsourcing data into the cloud, encryption measures are frequently employed to ensure secrecy, however, commercial storage providers are hesitant to utilize encryption owing to the possibility of various cipher texts. Convergent encryption has been suggested as a solution to these problems to protect data secrecy and permit duplication[6]. Systems for cloud storage have difficulty maintaining duplication, dependability, and secrecy. Duplication is a method for maximizing the use of storage resources by preventing the creation of duplicate copies of the same material. Single-cloud duplication architecture and multi-cloud duplication architecture are two categories in which it might be placed. Duplication techniques are being utilized more often to back up data and lessen network and storage transparency by minimizing data redundancy. For applications with high duplication ratios, such as archive storage systems, they are very helpful. File-level block-level duplication, client-side or server-side duplication, and cloud storage services like Google Drive and Drop Box can all be used as the foundation for duplication systems[10]. Duplication can, however, reduce system consistency, resulting in lost user data and unavailable data.

Reliability and data privacy are essential for ensuring high data consistency. Numerous advantages come with cloud storage, such as cost savings, accessibility, and scalability[4].

The difficulty for providers is to provide effective storage as data creation rates rise. duplication is a popular method employed by cloud storage companies that promise to save 90–95% of capacity. From secondary storage optimization to main storage and bigger storage spaces like cloud storage, this method has developed. Data owners frequently outsource their data in encrypted forms to cloud servers since they are outside of their perimeter of protection. Duplication is less practical due to encryption, which changes data into a cipher text format, which cannot be recognized[8].

Duplication and encryption are both necessary for data security and efficient storage, though. To achieve safe and efficient storage, duplication and encryption must cooperate. To determine their efficacy, several methods and techniques for duplication over encrypted data are investigated.However, in duplication systems shared access to user data might reveal data confidentiality flaws. Traditional encryption uses data similarity to identify redundancy, whereas duplication uses randomization services to produce cipher texts that are indistinguishable from one another.

Private data must be securely stored due to the exponential development in data quantities. While cloud computing has benefits like duplication, it also has issues with dynamic ownership management and access control[3].

II EXISTING SYSTEM

Duplication is a technique used by cloud storage companies like Drop box, Google Drive, and Mozy to store only one duplicate of each submitted file to conserve space[3]. However, because separate encryption keys are used when customers encrypt their data, storage savings are lost. Dude and other commercially available technologies for encrypted data duplication are ineffective. A hybrid cloud combines private and public clouds, storing some of the enterprise's most important data in its private cloud while making other data available through its public cloud[1].

Provable data possession (PDP) is a methodology that enables a client to confirm that the server has the original data without having to get it. By selecting random sets of blocks from the server, the model creates probabilistic proofs of possession while saving I/O expenses.

The current deduplication approaches still have issues with dynamic ownership management and access control[8]. The challenge/response protocol transmits a tiny, consistent quantity of data, minimizing network transmission, while the client keeps a constant amount of meta data to validate the evidence. In widely dispersed storage systems, the PDP paradigm for remote data verification accommodates enormous data sets. Even when compared to schemes that offer lesser guarantees, the two provably-secure PDP techniques that are given are more effective than earlier alternatives.

Rather than cryptography processing, disc I/O is what limits PDP performance.The PDP approach provides probabilistic proof that a third party is the file's storage location, enabling the server to access limited data from the file to produce the proof. Data deduplication makes it possible for data storage systems to identify and eliminate duplicate data without affecting the availability of that data[2]. The first remote data verification method that can be proven to be safe only stores a tiny amount of meta data and consumes bandwidth. The second, more effective variant, albeit it offers a less reliable assurance, uses a single modular exponentiation at the server to demonstrate data possession. Both strategies make use of homomorphic verifiable tags that can be merged to create a single value. A convergent key produced from the copy of the data is used to encrypt it, followed by a master key that is safely held locally by each user[6].

The foundation of an archive introspection system for the long-term preservation of astronomy data is the effective PDP scheme. The method is designed to preserve multi-terabyte astronomy data sets for the long term at a university library, where they will be copied at several locations. Freeloading occurs when partners seek to use the system's storage resources without giving back any of their own. A single pool of public cloud services from various independent vendors is combined by multi-cloud storage[4].

Each partner autonomously manages the location and physical implementation of these copies, and partners may even outsource storage to external servers.The quickest reaction time of any proof-of-retrieve ability approach with private variability is provided by the second strategy, which is based on pseudo-random functions (PRF). Both techniques combine a proof into a single tiny authentication value using homomorphic characteristics.

The PDP scheme's efficiency cannot be achieved by fetching whole file blocks to reduce computation and block accesses at the server. Deduplication is receiving more attention in academia and business due to its ability to increase storage efficiency and save space in high-deduplication applications [10]. Though the approaches also aim to reduce it, the client's computation complexity is less crucial. PDP techniques sample the server's storage by gaining access to a random portion of blocks to achieve performance objectives.

As opposed to a deterministic guarantee, which requires access to all blocks, this offers a probabilistic guarantee of possession. An MLE scheme is a type of symmetric encryption in which the message itself serves as the source of the key used for both encryption and decryption[5]. The client has the option of requesting evidence for each file block, making the assurance of data possession deterministic.

Sampling considerably improves the performance of demonstrating data possession since it uses a small number of blocks in the file to show data possession with a high degree of probability. Interestingly, regardless of the overall number of file blocks, the client may identify server misbehavior with a high probability when the server deletes a portion of the file by requesting evidence for a certain number of blocks. This paradigm is well-suited to our PDP solutions, which have little server overhead and little fluctuation in the amount of communication needed for each task.The homomorphic verifiable tags, which enable data ownership verification without requiring access to the actual data file, are essential elements of these approaches. Data storage is the most significant and well-liked cloud service.

Users of the cloud upload their data to the CSPs center and decide whether to keep it there[7].Existing research suggests outsourcing encrypted data to cloud storage service providers (CSP) to ensure data privacy. However, users might have encryption methods to protect duplicate data. Existing duplication techniques are vulnerable to brute-force assaults and have little flexibility in data access control and revocation. Rapid data development needs low-cost cloud storage options[9]. Although duplication technology is widely used in many different data storage situations, not all data centers may find it to be appropriate.

This E-Guide from Search-storage.com covers the advantages and disadvantages of DE-dupe backup as well as common misconceptions regarding deduce and compression on main storage.

DISADVANTAGES OF THE EXISTING SYSTEM

1. Without accessing every block, a deterministic guarantee cannot be given.

2. It doesn't guarantee that the data is yours.

3. It is unable to provide both consistence storage and sampling across blocks.

4.The scheme's security has not been established and is still in doubt.

5. Techniques for source authentication do not apply to the possession of proven data.

III PROPOSED SYSTEM

In this work, a unique server-side duplication approach for dynamic encrypted access is proposed. The plan makes use of safe ownership group key distribution and randomized convergent encryption to avoid data leakage to unauthorized users and trustworthy cloud storage services[8].

It strengthens security by ensuring data integrity against tag inconsistency attacks. The system has low computational overhead and is almost as efficient as earlier methods, according to the efficiency analysis results. The authors suggest Dekey, a novel architecture that eliminates the need for users to handle keys individually and instead securely distributes convergent key shares over several servers[6]. Dekey is shown to be secure through security analysis, and the Ramp secret sharing method serves as a proof-of-concept.

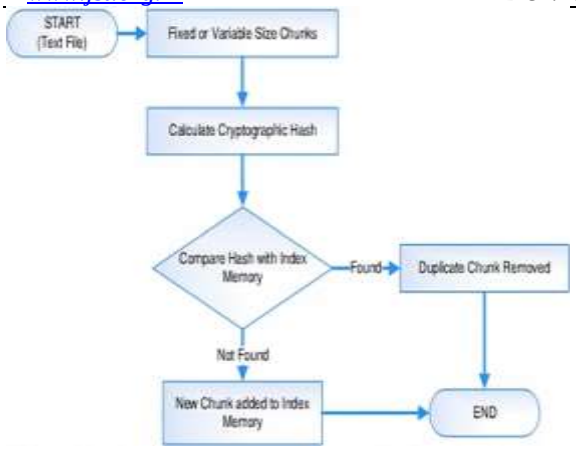

Fig.2 Process steps for deduplication

This work introduces the SPARK distributed duplication system, which tackles issues with tag stability and data privacy in cloud storage[9]. It outperforms earlier techniques in terms of security, efficacy, and applicability and substitutes a deterministic secret-sharing mechanism for convergent encryption. In a hybrid cloud architecture, where a private cloud handles duplication and dynamic ownership management and a public cloud manages storage, SPARK is a server-side duplication strategy for encrypted data.

This introduces R-MLE2 and introduces Cd-store, a unified multi-cloud storage solution for outsourcing backup data, formalizes Message-Locked Encryption (MLE) for safe duplication and space-efficient secure outsourced storage[2]. Convergent dispersion and two-stage duplication are combined in the CD store, which reduces bandwidth use and storage requirements while being resistant to side-channel assaults.

ADVANTAGES OF PROPOSED SYSTEM

1. The client is allowed to check for duplicate copies of selected records with the chosen subject.

2. Through the difficult process of encoding the record with unique permission keys, greater security is made possible.

3. Decrease the storage space of the tags for a reliability check. To strengthen the security of duplication and ensure data privacy.

4. The dictionary attack is avoided since no adversary could directly infer the convergent key from the file's content.

5. Any adversary lacking the file is unable to persuade the cloud storage provider to provide them with the necessary access privileges.

IV RESULT AND ANALYSIS

This section examines a scheme's security, paying particular attention to integrity, confidentiality, and defense against stub-reserved attacks. The CPABE scheme, message-locked encryption, symmetric encryption, convergent All or-nothing transform, and bloom filter are all considered to be secure underlying technologies.

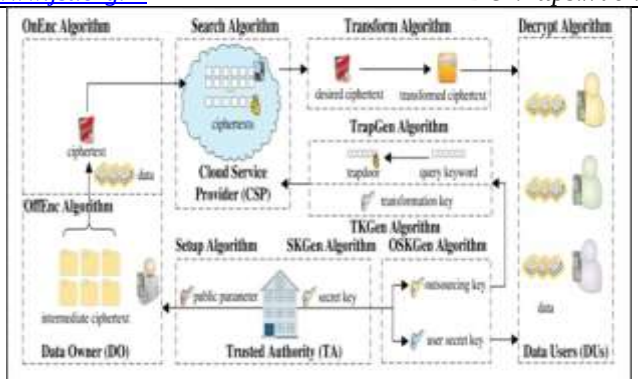

Fig.3.Final analysis

V CONCLUSION

In this paper, the idea of authorized data duplication was put forth in this project to safeguard data security by incorporating varied user privileges in the duplicate check.

Especially for data operations, encrypted data management with duplication is essential for safe cloud storage services. Future work should concentrate on establishing a customization solution for duplication and data access controlled by the data owner or representative agent, scheme optimization with hardware acceleration for IOT devices, and fast data ownership verification.

VI REFERENCESS

1) J. Li, Y. K. Li, X. Chen, P. P. C. Lee, and W. Lou, "A hybrid cloud approach for secure authorized deduplication," IEEE Trans. Parallel Distrib. Syst., vol. 26, no. 5, pp. 1206–1216, 2015.

2) T. Jiang, X. Chen, Q. Wu, J. Ma, W. Susio, and W. Lou, "Secure and efficient cloud data deduplication with randomized tag," IEEE Trans. Information Forensics and Security, vol. 12, no. 3, pp. 532–543, 2017.

3) J. Hur, D. Koo, Y. Shin, and K. Kang, "Secure data deduplication with dynamic ownership management in cloud storage," IEEE Trans. Knowl. Data Eng., vol. 28, no. 11, pp. 3113–3125, 2016.

4) M. Li, C. Qin, and P. P. C. Lee, "Cdstore: Toward reliable, secure, and cost-efficient cloud storage via convergent dispersal," in 2015 USENIX Annual Technical Conference, USENIX ATC '15, 2015, pp. 111–124.

5) B. Mihir, K. Sriram, and R. Thomas, "Message-locked encryption and secure deduplication," Advances in Cryptology - EUROCRYPT 2013, 32nd Annual International Conference on the Theory and Applications of Cryptographic Techniques,, vol. 28, no. 11, pp. 296–312, 2013.

6) J. Li, X. Chen, M. Li, J. Li, P. P. Lee, and W. Lou, "Secure deduplication with efficient and reliable convergent key management," IEEE trans. Parallel Distrib. Syst., vol. 25, no. 6, pp. 1615–1625, 2014.

7) Z. Yan, W. Ding, X. Yu, H. Zhu, and R. H. Deng, "Deduplication on encrypted big data in cloud," IEEE Trans. Big Data, vol. 2, no. 2, pp. 138–150, 2016.

8) S. Jiang, T. Jiang and L. Wang, "Secure and Efficient Cloud Data Deduplication with Ownership Management," in IEEE Transactions on Services Computing, vol. 13, no. 6, pp. 1152-1165, 1 Nov.-Dec. 2020, doi: 10.1109/TSC.2017.2771280.

9) Dave, J., Faruki, P., Laxmi, V., Zemmari, A., Gaur, M., & Conti, M. (2020). SPARK: Secure Pseudorandom Key-based Encryption for Deduplicated Storage. Computer Communications, 154, 148-159. https://doi.org/10.1016/j.comcom.2020.02.037.

10) J. Li et al., "Secure Distributed Deduplication Systems with Improved Reliability," in IEEE Transactions on Computers, vol. 64, no. 12, pp. 3569-3579, 1 Dec. 2015, doi: 10.1109/TC.2015.24*01017.*