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 13, Issue 11, November 2024

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.524

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

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

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 11, November 2024

|DOI: 10.15680/IJIRSET.2024.1311126|

A Survey on Efficient Data Storage Management and Reducing Deduplication in Cloud Computing

Mrs. P. N. Wadibhasame, Khutwad Rutuja Sunil, Ranjana Krishnalal Yadav,

Kashish Kalpeshkumar Katariya, Akanshka Chimanlal Kumbhar

Department of Computer Engineering, Sinhgad Academy of Engineering, Kondhwa, Maharashtra, India

ABSTRACT: Cloud storage providers can save a lot of storage and bandwidth by adopting data deduplication, which eliminates duplicated data by keeping only one copy of a file. Cloud storage services enable people and businesses to outsource data storage to remote servers. Deduplication is becoming a widely utilized technology in cloud data centres to improve the efficiency of IT resources. The duplication check allows an attacker to determine whether a file (such as a pay stub containing a name and salary amount) has already been stored (by someone else), exposing the user's private data. Because of reasons like as high computational cost, low attack resistance, information leakage, and others, the bulk of previous techniques that rely on the help of a trustworthy key server (KS) are insecure and restricted. A single point of failure occurs when the entire system fails due to the failure of a trustworthy KS. Data duplication in the current system causes security issues. As a result, we are adopting double encryption and hash code generation to tackle these issues. It combines access control with cloud data deduplication. Only authorized data holders have access to the symmetric keys required for data decryption, thus encrypted data can be accessed safely. Our technique is efficient and safe within the stated security model, and it is ideal for enormous data deduplication, according to a thorough performance analysis and testing. We employ computer simulations and in-depth analysis to evaluate its performance. The results show that the approach is more effective and efficient for real-world applications, notably for large-scale data deduplication in cloud storage.

KEYWORDS: Storage Management, cloud computing, deduplication, Manage Storage.

I. INTRODUCTION

Cloud computing reorganises different Internet resources to provide a new manner of offering services. Data storage is one of the key and most often used cloud services. Usually, data is stored encrypted to protect the privacy of the data subjects. Encrypted data, however, poses significant difficulties for cloud data deduplication, which is essential for processing and preserving vast amounts of data. Conventional deduplication techniques cannot be applied to encrypted data. Data deduplication suffers from inadequate security in current encryption techniques. They are not resilient to being in charge of and denying access to data. Consequently, a few of them. They are simple to put into practise. In this paper, we suggest rewriting the proxy and using challenge features to deduplicate encrypted data that is stored in the cloud. Access control and cloud data deduplication integrated. We use computer simulations and in-depth analytics to determine how well they performed. The results demonstrate the best possible efficiency and efficiency scheme distribution in practise, particularly for large-scale cloud data deduplication storage.

II. METHODOLOGY

The integrity of uploaded files in the CSP with deduplication is checked using a unique protocol called TDICP based on PIR. While the verification tags can be copied at the CSP, TDICP enables users to create their own unique verification tags for integrity check. To ensure the accuracy of the PSI-based duplication check and prevent the CSP from tricking the user into paying for storage space that isn't being used because of deduplication, we propose the UDDCP protocol. We create VeriDedup, a brand-new deduplication technique, which includes the two innovative protocols mentioned above along with other crucial characteristics like POW and data.

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

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 11, November 2024

|DOI: 10.15680/IJIRSET.2024.1311126|

To solve these issues, VeriDedup is a verifiable deduplication scheme. It is able to provide a guarantee regarding the accuracy of the duplication check and to provide configurable tag creation for integrity check over encrypted data deduplication in an integrated manner. We specifically recommend a innovative Private Information Retrieval (PIR)-based Tag-flexible Deduplication-supported Integrity Check Protocol (TDICP) by presenting the note set verification tag, a revolutionary verification tag that enables numerous users to generate their own unique verification tags while sharing the same file. tag deduplication at the CSP and tag verification. Additionally, we try our best to ensure the accuracy of introducing a new User Determined Duplication Check Protocol (UDDCP) based on Private Set Intersection to check for data duplication (PSI), which can prevent a CSP from producing a fictitious duplication check result.

III. LITERATURE SURVEY

1. Information One method for lowering the quantity of storage space required by an organisation to store its data is de-duplication. Many data items are duplicate copies in the storage systems of the majority of organisations. For instance, a same file might be saved by multiple users in various locations, or a large amount of the same information could be present in two or more non-identical files. By preserving only one copy of the data and substituting pointers that point back to the original copy for the other copies, de-duplication gets rid of these superfluous copies. De-duplication is often used by businesses for disaster recovery and backup purposes, but it can also be utilised to free up space in primary storage [1].
2. One of the key data compression strategies for removing redundant copies of data is data de-duplication, which is extensively utilised in cloud storage to conserve bandwidth and decrease storage space. To safeguard sensitive data's privacy while encouraging de-duplication [2].
3. This work demonstrates that numerous methods are employed to remove redundant copies of data that repeat themselves; among these methods, data duplication is a crucial data compression method. There are several benefits to data duplication, the primary one being that it will use less bandwidth and less storage space when used for cloud storage. In order to safeguard the privacy of sensitive information and facilitate deduplication, data is encrypted prior to outsourcing using the suggested convergent encryption technique. The initial draught of this document formally addressed issues with authorised data duplication for improved data security protection [3].
4. A de-duplication method for storing private data is shown and codified. It makes sense that a private data deduplication protocol would enable a client to affirm to a server that they are the owners of a summary string of the data that they possess the private data without disclosing any additional information to the server. Our idea can be seen as an enhancement to the most advanced public data de-duplication protocol available. In the context of two-party computations, the security of private data de-duplication methods is formalised in the simulation-based framework. Next, a presentation and analysis of private de-duplication techniques constructed using common cryptographic assumptions follow. We demonstrate that, under the assumptions that the underlying hash function is collision-resilient, the discrete logarithm is hard, and the erasure coding algorithm can erase up to a fraction of the bits in the presence of hostile adversaries, the suggested private data deduplication protocol is provably secure [4].

IV. EXISTING SYSTEM APPROACH

Brute-force attacks plague deduplication methods currently in use. They are unable to cooperatively provide both revocation and control of data access simultaneously. The majority of current solutions are unable to combine good performance with security, privacy, and dependability. There may be a storage delay if the first data holders aren't always online or accessible for each management. For data holders to be involved in the deduplication process, second deduplication may become too complex in terms of computation and communication. Third, while looking for duplicate data, it might violate the privacy of the data owner. Due to data suffer dispersion, a data holder may not know how to provide users with a deduplication key or data access right in some situations where it does not know other data holders.

About the ownership dilemma and PRE. Even when the data holders are not online, our scheme can facilitate data sharing and updates in a flexible manner with deduplication. Because only authorised data holders have access to the symmetric keys needed for data decryption, encrypted data can be securely accessed. A thorough performance analysis and test revealed that, in the context of the given security model, our technique is both secure and effective, making it ideal for deduplication. The outcome of our computer simulations demonstrated our scheme's practicality even further.

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

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 11, November 2024

|DOI: 10.15680/IJIRSET.2024.1311126|

V. PROPOSED SYSTEM APPROACH

There have been numerous issues with cloud storage spaces in recent times. Memory is wasted if the data owner stores a file in the cloud that is already there.

Thus, we do away with the duplication in this publication. Additionally, security problems arise while storing data on cloud servers. With the use of hash code algorithms to generate hashe values and double encryption techniques, we are resolving these issues.

Data Holders

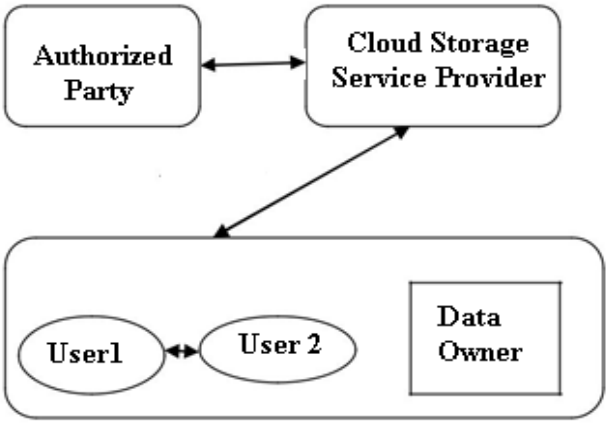

Fig 1: System architecture

CSP: The CSP permits data owners to use its storage services. It is not entirely reliable. As a result, it is interested in the information that is saved. In order to make money, it should store data honestly.

Data Holder: The data owner has the ability to upload and save files and data in the CSP. Several data holders may store their files in the CSP as encrypted raw data using this technique. The person who develops or creates the file is considered the data owner. The data holder is more normal than the owner, who has a higher priority.

AP: an approved party (AP) in whom the data holders have complete faith. The data holders to manage data deduplication and confirm data ownership. It doesn't work in concert with CSP. In this instance, CSP shouldn't be aware of the simple user data that is stored there. AP is unable to access the data kept in CSP.

VI. CONCLUSION

Discuss how to successfully manage encrypted data with deduplication in order to achieve cloud storage services for large data sets. This study proposes a deduplication-based solution to managing encrypted huge data in the cloud, which is based on the ownership challenge and PRE. Even when data holders are offline, this technique allows for flexible and deduplicative data sharing and changes. The symmetric keys required for data decryption by approved data holders allow for secure access to encrypted data. According to the big data deduplication security paradigm, the method is secure, as evidenced by the performance analysis and graphs. The graph's variation also showed the results of the computer simulations.

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

Impact Factor 8.524

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 13, Issue 11, November 2024

|DOI: 10.15680/IJIRSET.2024.1311126|

REFERENCES

[1] M. Bellare, S. Keelveedhi, and T. Ristenpart, "DupLESS: Server aided encryption for deduplicated storage," in Proc. 22nd USENIX Conf. Secur., 2013, pp. 179–194.
[2] J. Li, Y. K. Li, X. F. Chen, P. P. C. Lee, and W. J. Lou, "A hybrid cloud approach for secure authorized deduplication," IEEE Trans. Parallel Distrib. Syst., vol. 26, no. 5, pp. 1206–1216, May 2015, doi:10.1109/TPDS.2014.2318320.
[3] P. Meye, P. Raipin, F. Tronel, and E. Anceaume, "A secure two phase data deduplication scheme," in Proc. HPCC/CSS/ICESS, 2014, pp. 802–809, doi:10.1109/HPCC.2014.134.
[4] Z. Yan, X. Y. Li, M. J. Wang, and A. V. Vasilakos, "Flexible data access control based on trust and reputation in cloud computing," IEEE Trans. Cloud Comput., vol. PP, no. 99, Aug. 2015, doi:10.1109/ TCC.2015.2469662, Art.
[5] P. Puzio, R. Molva, M. Onen, and S. Loureiro, "ClouDedup: Secure deduplication with encrypted data for cloud storage," in Proc. IEEE Int. Cof. Cloud Comput. Technol. Sci., 2013, pp. 363–370, doi:10.1109/CloudCom.2013.54.
[6] C. Y. Liu, X. J. Liu, and L. Wan, "Policy-based deduplication in secure cloud storage," in Proc. Trustworthy Comput. Serv., 2013, pp. 250–262, doi:10.1007/978-3-642-35795-4_32.
[7] Z. Sun, J. Shen, and J. M. Yong, "DeDu: Building a deduplication storage system over cloud computing," in Proc. IEEE Int. Conf. Comput. Supported Cooperative Work Des., 2011, pp.
[8] Z. C. Wen, J. M. Luo, H. J. Chen, J. X. Meng, X. Li, and J. Li, "A verifiable data deduplication scheme in cloud computing," in Proc. Int. Conf. Intell. Netw.Collaborative Syst., 2014, pp. 85–90, doi:10.1109/INCoS.2014

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

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