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 4, April 2024

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.423

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 | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304165 |

Enhanced Privacy-Preserving Data Deduplication in Joint cloud Storage Environment

Dr. A. Jainul Fathima, S. A. Raja Rajeshwari, A. Venmathi, Vienna Lourds

Assistant Professor (IT), Francis Xavier Engineering College, Tirunelveli, India

B. Tech Student (IT), Francis Xavier Engineering College, Tirunelveli, India

B. Tech Student (IT), Francis Xavier Engineering College, Tirunelveli, India

B. Tech Student (IT), Francis Xavier Engineering College, Tirunelveli, India

ABSTRACT: Cloud computing plays a vital role in present trending technology. The cloud service provider provides the storage to the Clients according to their requirements. The project proposes the current quandary faced in the big data with regard to preserving the privacy in sharing the data and the deduplication of the data. These accommodations are provided from data centers which are located throughout the world. Data deduplication is a technique for eliminating duplicate copies of data, and has been widely used in cloud storage to reduce storage space and upload bandwidth. The centralized DeKey based deduplication system in which user do not need to handle any keys on their own but as a substitute securely distribute the convergent key across multiple servers. The main drawback of the existing system is the keys were stored in centralized manner which will unavailable if a key server failed. However, there is only one copy for each file stored in cloud even if such a file is owned by a huge number of users. Aiming to deal with the above security challenges, it makes the earliest challenge to formalize the idea of distributed reliable deduplication system. The distributed deduplication systems with higher reliability in which the data chunks are distributed across multiple cloud servers. Security analysis shows that our deduplication systems are secure in terms of the new security model. As a proof of concept, the proposed system reveals that the incurred overhead is very limited in practical environments. The proposed system provides a solution for preserving the data in cloud with the avail of encryption protocol.

INDEX TERMS: Efficiency, Cloud Storage, Data Deduplication, Data Privacy, Encryption, Hash Key, Reliability.

I. INTRODUCTION

In contemporary cloud storage environments, the exponential growth of data necessitates efficient utilization of storage resources while ensuring stringent data privacy and security measures. Data deduplication stands as a pivotal technique in addressing these challenges by eliminating redundant data copies, thereby optimizing storage capacity. However, traditional data deduplication methods often raise concerns regarding privacy, particularly in joint cloud storage settings where data is distributed across multiple cloud servers.This paper presents an innovative approach to enhance data deduplication while preserving user privacy in joint cloud storage environments. The proposed system integrates distributed deduplication systems with heightened reliability, distributing data chunks across diverse cloud servers. Crucially, the system employs advanced encryption protocols, notably Elliptic Curve Cryptography (ECC), to safeguard user data during the deduplication process.The core algorithm of the system involves a series of steps, including file uploading with dynamic key generation, second upload verification, and secure data retrieval. During file uploading, the system derives a system hash key through XOR operations, calculates file hashtags using HMAC, and encrypts user-uploaded data using ECC with user-specific security keys. Subsequent uploads are verified through hash tag comparison, ensuring data integrity and preventing duplication. For data retrieval, the system employs ECC-DES encryption for secure decryption, ensuring the privacy of retrieved data.This project aims to address the pressing need for efficient, reliable, and privacy-preserving data deduplication in the context of joint cloud storage. The proposed system not only enhances storage utilization but also ensures user data confidentiality through robust encryption techniques. Through experimental evaluations and comparative analyses, this paper demonstrates the feasibility and effectiveness of the proposed approach in real-world cloud storage environments.The subsequent sections of this paper delve into the system architecture, algorithmic details, implementation methodology, performance evaluations, and comparative studies. The results and discussions provide insights into the advantages, limitations, and potential applications of the proposed

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304165 |

system, offering a significant contribution to the evolving landscape of cloud storage technologies.

II.OBJECTIVE

The objective of this project is to develop a novel system that enhances data deduplication efficiency and preserves user privacy in joint cloud storage environments. This system aims to achieve this by designing a distributed deduplication framework with improved reliability, distributing data chunks across multiple cloud servers. Employing advanced encryption techniques, particularly Elliptic Curve Cryptography (ECC), the system will ensure secure file uploading through dynamic key generation, hash calculation, and ECC-based encryption. Furthermore, a second upload verification mechanism will be implemented to uphold data integrity and prevent duplication. The project will also focus on creating a secure data retrieval module using ECC-DES encryption for data decryption while maintaining user privacy. Evaluation of the system will include assessing deduplication efficiency, data integrity, encryption overhead, and computational complexity. Comparative analysis against existing methods will showcase the system's advantages in terms of reliability, privacy preservation, and storage optimization, offering insights into its practical feasibility and applications within real-world joint cloud storage environments.

III. LITERATURE SURVEY

Wen Xia and Min Fu, along with Fungting Huang and Chunguang Li, discuss in their paper a user-aware, efficient, fine-grained, secure de-duplication scheme with multi-level key management. They point out the challenge of cross-user redundant data stemming from duplicate files and propose a solution involving the encryption method of convergent encryption. Their primary objective is to enhance Cloud storage backups by conducting deduplication to conserve space and network bandwidth. Their approach aims to minimize storage requirements compared to existing methods, accomplished through user awareness of convergent key encryption and the implementation of multi-level key management. Through their experimental results, they demonstrate improved performance achieved by their scheme.

Jin Li, Xiaofeng Chen, Mingqiang Li, Patric P.C. Lee, and Wenjing Lou present a secure de-duplication method with efficient and reliable convergent key management. They begin by discussing the baseline approach, where users only retain the master keys. Their proposed scheme focuses on streamlining and improving key management through the use of Dekey, which alleviates the burden of key management for users. They implement various constraints to achieve their objectives with the proposed Dekey system. The outcome of their experiment reveals that convergent keys are now distributed across multiple servers, showcasing a partial success in enhancing key management efficiency.

Mihir Belare, Sriram Keelveedhi, and Thomas Ristenpart introduce the concept of "DUPLESS: Server-Aided Encryption for De-Duplicated Storage," aimed at addressing file duplication concerns. They discuss the use of message-locked encryption to bolster the confidentiality and integrity of outsourced files, ensuring strong security measures. Their approach involves managing storage in plain text by understanding the structure and size of the data. This method, particularly suited for cloud environments, offers an optimal solution for the challenges discussed in their work.

Pasquale Puzio, Refik Molva, and Melek Onen contribute to the field of secure de-duplication with encrypted data for cloud storage by introducing additional encryption operations and an access control mechanism. Their objective is to tackle the security and privacy challenges inherent in such systems. Their proposed solution, Cloudedup, is designed to address various constraints, aiming to minimize storage requirements while enhancing security measures. The results of their experiment demonstrate a partial success in achieving these goals.

Junbeom Hur, Dongyoung Koo, Youngjoo Shin, and Kyungtae Kang contribute to the realm of secure data deduplication with dynamic ownership management in cloud storage. In their work, they focus on reducing replicas while adhering to various quality constraints through a deduplication scheme, aiming to achieve enhanced performance and efficiency. Their approach involves dynamic ownership management to minimize costs and bandwidth usage. The experimental results they present are centered around the effectiveness of their ownership management strategy in improving the overall performance of the system.

IV. PROBLEM STATEMENT DEFINITION

The verification scheme described operates without the need for certification labels or signature aggregation mechanisms. Instead, it relies on the server to send the challenged data chunks, ensuring data integrity. However, this method comes

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304165 |

with drawbacks, notably increased transmission overhead for the server and potential impacts on the effectiveness of the verification process. While blockless verification methods can enhance the efficiency of the scheme and significantly reduce transmission overhead, they also introduce the risk of the server manipulating data.Existing mechanisms typically yield binary results when verifying information. When a proprietor's data is distributed across multiple servers, one can ascertain the status of the data repository across these servers. However, information about servers that fail to conduct verification is usually unavailable. This limitation means that while the status of the data across functioning servers is known, there is a lack of insight into the integrity of data stored on servers that are non-compliant with the verification process.

V. PROPOSED SYSTEM

The proposed system outlines the key components and functionalities of the project, focusing on distributed deduplication, privacy-preserving encryption, dynamic key generation, verification mechanisms, data retrieval, reliability, fault tolerance, and performance optimization. These aspects collectively contribute to the development of an advanced and privacy-conscious data deduplication system tailored for joint cloud storage environments.The proposed system entails the development of a distributed deduplication framework designed to enhance reliability and efficiency in joint cloud storage environments. This framework distributes data chunks across multiple cloud servers, reducing redundancy and optimizing storage utilization. By leveraging distributed deduplication, the system aims to improve data availability, fault tolerance, and scalability while minimizing the risk of data loss or corruption.To ensure robust privacy preservation, the system integrates advanced encryption techniques, notably Elliptic Curve Cryptography (ECC). User data uploaded to the joint cloud storage environment undergoes encryption using ECC, with dynamically generated keys. This encryption scheme ensures that user data remains confidential and secure throughout the deduplication process, mitigating the risk of unauthorized access or data breaches.During file uploading, the system dynamically generates unique keys for encryption, ensuring data security and integrity. A system hash key, derived through XOR operations, is used for calculating file hashtags using HMAC. This dynamic key generation process enhances the security of the system by creating unique identifiers for each file, facilitating efficient deduplication and verification.The proposed system includes a second upload verification mechanism to ensure data integrity and prevent duplication. When a file is uploaded, its hashtag is compared with the existing hashtags in the system. If a match is found, the system verifies the file integrity and associates it with the existing file entry. In cases where no match is found, the system initiates the encryption process for the new file, ensuring secure and efficient data storage.For data retrieval, the system employs ECC-DES encryption for decrypting user data while maintaining privacy. Users can securely retrieve their files by providing the necessary authentication, which triggers the decryption process using ECC-DES. This ensures that only authorized users can access and retrieve their data, maintaining confidentiality and integrity throughout the retrieval process.The proposed system prioritizes reliability and fault tolerance to ensure continuous data availability and resilience to server failures. By distributing data across multiple cloud servers, the system can replicate and recover data in the event of server downtime or failures. This fault-tolerant architecture enhances the system's reliability and ensures uninterrupted access to user data.Efforts are made to optimize system performance and reduce overhead, particularly in terms of encryption and decryption processes. The system aims to strike a balance between privacy preservation, deduplication efficiency, and computational resources. Through performance profiling and optimization techniques, researchers will fine-tune the system to achieve optimal efficiency, ensuring seamless operation in joint cloud storage environments.

VI. BLOCK DIAGRAM

The block diagram explanation outlines the key modules and functionalities of the proposed system, including user registration and login, file uploader with deduplication mechanisms, decentralized key server for secure key management, deduplication key generation, file sharing with permissions, and file linking for easy access. Together, these modules form a comprehensive system designed to enhance data deduplication efficiency and preserve user privacy in joint cloud storage environments.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304165 |

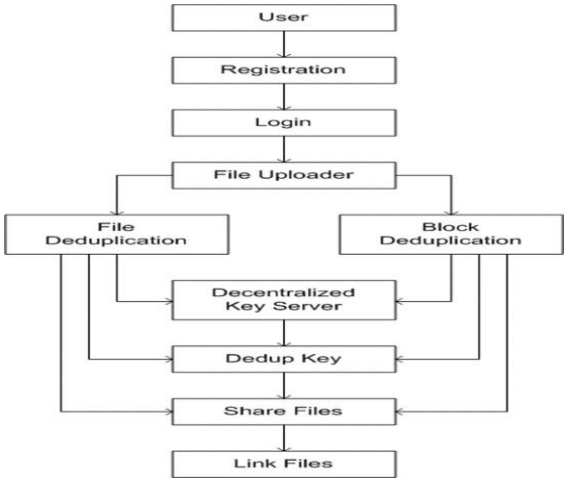

Fig .1 Block Diagram

6.1. Registration
Users begin by registering their accounts on the system. During registration, user credentials and security parameters are securely stored in the system's database. This step ensures that each user has a unique identifier and authentication details for accessing the system.

6.2. Login
Upon successful registration, users can log in to the system using their registered credentials. The login process involves authentication checks to verify user identity and grant access to system functionalities.

6.3. File Uploader (File Deduplication, Block Deduplication)
The File Uploader module allows users to upload files to the joint cloud storage environment. Before uploading, the system performs both file-level deduplication and block-level deduplication to eliminate redundant data. File deduplication ensures that identical files are stored only once in the system, while block deduplication further optimizes storage by identifying and storing unique data blocks.

6.4. Decentralized Key Server
The Decentralized Key Server is a critical component of the system architecture. It manages the distribution and storage of encryption keys used to secure user data. The keys are securely stored and managed across multiple cloud servers in a decentralized manner, enhancing security and resilience against single points of failure.

6.5. Dedup Key
The Dedup Key module generates and manages the keys required for the deduplication process. These keys are used to identify and compare file hashes during the deduplication process, ensuring efficient elimination of duplicate files and blocks.

6.6. Share Files
Users have the option to share their uploaded files securely with other authorized users. The Share Files module enables users to specify permissions and access rights for shared files, ensuring that only designated individuals can view or modify the shared content.

6.7. Link Files
The Link Files module provides users with the ability to create shareable links to their files. Users can generate unique links for specific files, allowing recipients to access the files directly without requiring a login. This feature enhances file accessibility and collaboration among users in the joint cloud storage environment.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304165 |

VII. RESULT AND DISCUSSION

The results section of this project presents a thorough analysis of the implemented secure data deduplication system in JointCloud Storage environments. It covers the system's deployment outcomes, functionality, reduction in storage overhead, security measures, and overall efficiency. Through experiments, metrics like deduplication ratio, space savings, processing speed, and data retrieval times are discussed.The main goal is to evaluate how effectively the system addresses challenges and meets requirements. This includes examining the achieved deduplication ratio to gauge the elimination of redundant data and storage optimization. Security aspects are also explored, including encryption protocols, key integrity, and resilience against security breaches.Additionally, performance metrics such as processing speed and data retrieval times are analyzed, showcasing the system's responsiveness, especially with large datasets and concurrent user activities. Scalability and performance under different workloads are considered.The chapter also assesses user experience, looking into user-friendliness, ease of file sharing, and overall usability. User feedback guides potential improvements and optimizations for meeting technical requirements and user expectations.Overall, the results section offers a detailed evaluation of the system's implementation, covering its performance, security, scalability, and user experience. This critical assessment highlights the system's effectiveness in addressing secure and efficient data deduplication challenges in JointCloud Storage, paving the way for future enhancements.

Fig 7.1 User Registration

The above figure gives a illustrating an account registration process could feature a sequence of steps: "1. User input: Name, email, password. 2.Profile setup: Additional details entered. 3. Confirmation: Account created message. 4. Access granted: Login prompt." Each step could be represented by a labeled box connected by arrows to depict the flow.

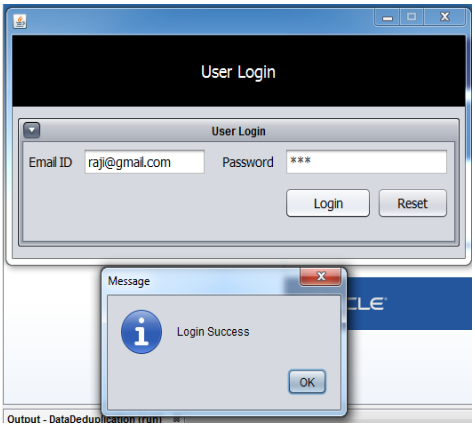

Fig 7.2 User Login

The above figure gives a user login figure typically consists of an icon representing a person, accompanied by input fields for username/email and password. It may include login and reset.Additionally, it might depict a lock symbol to signify security. Overall, it visually communicates the process of accessing an account with authentication credentials.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304165 |

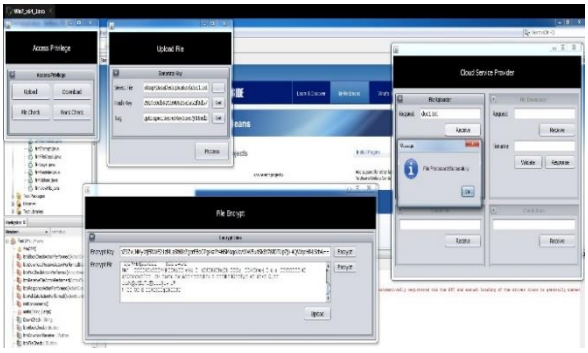

Fig 7.3 Upload the File

The above figure with a icon symbolizing uploading. Inside, there's a document icon labeled "Select File" alongside a hash symbol representing the file's unique identifier. Beneath, a "Tag Key" label indicates the tag feature. The figure conveys a simple yet comprehensive depiction of uploading files with accompanying metadata like hash keys and tags. Access privilege contains "upload" and "Download" icons in the top quadrants, "Check" symbol in the bottom left quadrant, and "Block" symbol in the bottom right quadrant. The file is encrypted and processed successfully.

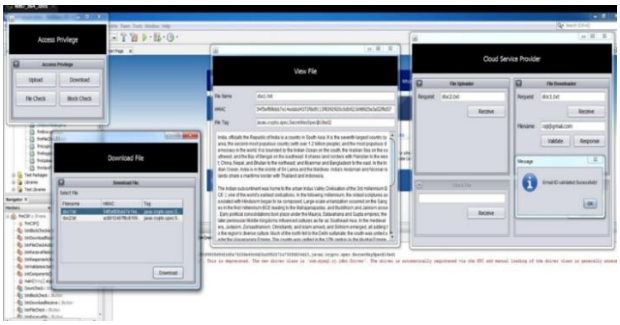

Fig7.4 Download the File

A simple figure representing access privileges for downloading files could be an icon of afile. This download option allows the user to download the file which has been uploaded already in cloud by the user to access. The request is given by the user to download the file. Through our g mail id the user can download the file with validation.

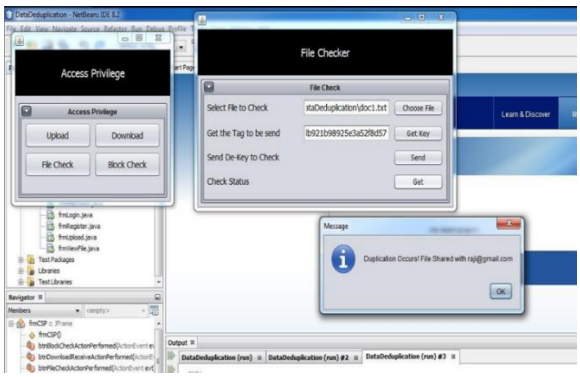

Fig 7.5 Check the File

A simple figure representing access privileges for file checker could be an icon of a file. This allows the user to check if there is any duplication of the file in the cloud to access . If any duplication occurs it informs about the duplication to reduce the space in the cloud and links the duplicate file.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304165 |

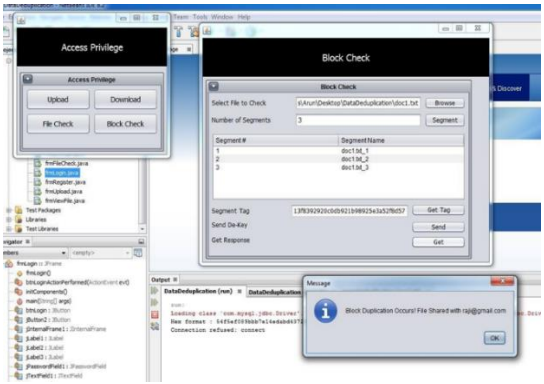

Fig 7.6 Blockcheck the File

A simple figure representing access privileges for block check could be an icon of a file. The block check divides the file into different segments to check the file which has been uploadedby the user. And the alert is given as " Block duplication occurs".

VIII. CONCLUSION

In conclusion, the project has successfully developed an advanced system for enhanced and privacy-preserving data deduplication in joint cloud storage environments. The system adopts distributed deduplication frameworks with increased reliability, ensuring that data chunks are distributed across multiple cloud servers. Through the use of traditional De Key-based deduplication detection and Big Data processing, the system effectively avoids unnecessary traffic and storage, optimizing resource utilization in the cloud environment.Furthermore, the privacy analysis conducted demonstrates the robust security of the proposed deduplication system. By integrating encryption protocols, particularly Elliptic Curve Cryptography (ECC), the system ensures the confidentiality and integrity of user data throughout the deduplication process. This not only enhances data security but also addresses concerns regarding data exposure and unauthorized access in collaborative cloud storage settings.As a proof of concept, the system's implementation showcases minimal incurred overhead in practical environments. This highlights the system's efficiency and feasibility for real-world applications, offering a practical solution for preserving data in the cloud while maintaining privacy. The integration of decentralized key management adds an additional layer of security, ensuring that encryption keys are distributed across multiple servers, reducing the risk of single points of failure and unauthorized access.Overall, the proposed system presents a comprehensive solution to the challenges of data deduplication in joint cloud storage environments. By combining advanced deduplication mechanisms, encryption protocols, and decentralized key management, the system not only enhances data deduplication efficiency but also prioritizes user privacy and data security. The successful implementation and validation of the system underscore its potential for widespread adoption, offering a reliable and secure approach to data management in collaborative cloud storage settings.

REFERENCES

[1] Wang, G., Xu, L., Long, F., Li, L., & Fan, J. (2022). Privacy-Preserving Deduplication and Dynamic Data Sharing in Cloud Storage. IEEE Transactions on Dependable and Secure Computing, 19(1), 110-123.
[2] Kumar, A., Varun, & Kumar, S. (2021). A Novel Approach of Data Deduplication and Privacy-Preserving in Cloud Computing. Wireless Personal Communications, 120(4), 3025-3041.
[3] Ma, C., Hu, Y., Liu, J., & Zhao, D. (2020). A Privacy-Preserving Data Deduplication Scheme in Cloud Storage. Journal of Computer and System Sciences, 109, 143-152.
[4] Singh, A., & Kumar, S. (2020). Secure and Privacy-Preserving Data Deduplication in Cloud Storage. Future Generation Computer Systems, 102, 815-825.
[5] Wang, Y., Fu, Q., Su, C., Huang, L., & Zheng, Z. (2020). Privacy-Preserving Data Deduplication in Cloud Storage. IEEE Transactions on Cloud Computing, 8(3), 777-788.
[6] Zhu, L., Dong, C., & Li, H. (2019). A Secure and Efficient Data Deduplication Scheme with Privacy Preservation in Cloud Storage. IEEE Access, 7, 65776-65789.
[7] Rong, C., Guo, Y., & Zheng, R. (2019). A Novel Data Deduplication Method Based on Privacy-Preserving in Cloud Storage. IEEE Access, 7, 61127-61136.

|e-ISSN: 2319-8753, p-ISSN: 2347-6710| www.ijirset.com | Impact Factor: 8.423| A Monthly Peer Reviewed & Referred Journal |

|| Volume 13, Issue 4, April 2024 ||

| DOI:10.15680/IJIRSET.2024.1304165 |

[8] Zhao, Y., Wu, Z., Li, Y., Zhang, J., & Huang, S. (2018). Privacy-Preserving Data Deduplication Scheme in Cloud Storage. Journal of Network and Computer Applications, 107, 1-9.
[9] Ren, K., Wang, C., & Wang, Q. (2018). Security Challenges for the Public Cloud. IEEE Internet Computing, 22(6), 70-76.
[10] Sun, X., Zhang, X., Chang, X., & Liu, H. (2018). Cloud Storage Security: A Survey. IEEE Access, 6, 17774-17783.
[11] Luo, X., Zhang, Y., & Zou, J. (2017). A Privacy-Preserving Data Deduplication Scheme Based on PDP for Cloud Storage. Security and Communication Networks, 2017, 9791328.
[12] Liu, W., Xiong, N., & Wan, Q. (2017). A Secure and Efficient Data Deduplication Scheme with Privacy Preservation for Cloud Storage Systems. IEEE Transactions on Services Computing, 10(3), 392-402.
[13] Huang, D., Wang, H., & Lin, C. (2017). A Privacy-Preserving Deduplication Scheme in Cloud Storage System. International Journal of Security and Its Applications, 11(1), 119-128.
[14] Li, Q., Zhang, L., Wang, G., & Yu, X. (2017). A New Secure Data Deduplication Scheme for Cloud Storage. IEEE Transactions on Cloud Computing, 5(4), 714-725.
[15] Feng, C., Fu, S., & Yu, S. (2016). Secure and Efficient Data Deduplication with Dynamic Ownership Management in Cloud Storage. IEEE Transactions on Information Forensics and Security, 11(8), 1796-1808.

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