International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 03 | March - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

DedupShield: Secure Cloud Storage Optimization

Dr. Harika.B[1], Chitta Srinivas[2], Gangam Rithvik Reddy[3]
[1] *Associate Professor, Mahatma Gandhi Institute of Technology*
[2,3] *UG Student, Mahatma Gandhi Institute of Technology*

Abstract- **In the digital age, managing the exponential growth of textual data while ensuring security in cloud environments is a significant challenge. Dedup Shield introduces a secure data deduplication system utilizing Advanced Encryption Standard (AES) for both encryption and decryption processes, aimed at enhancing storage efficiency and data security. By leveraging AES, the system ensures that textual data stored in the cloud remains confidential and protected from unauthorized access. The proposed framework delineates specific roles for cloud administrators, users, owners, and potential attackers, each with clearly defined functionalities. For instance, owners can securely upload and manage files, authorize user access, and perform file audits, while users can request access to these files under stringent security protocols. The SQL database supports these operations, maintaining the integrity and availability of the data. Studies indicate that data deduplication can reduce storage needs to great extent, highlighting the effectiveness of such systems in managing large-scale data efficiently while ensuring robust security measures .This approach not only enhances storage efficiency but also significantly mitigates the risks associated with data breaches, making Dedup Shield**

Keywords: **Digital age, textual data, security, cloud environments, Dedup Shield, data deduplication, Advanced Encryption Standard (AES), encryption, decryption, storage efficiency, data security, confidentiality, unauthorized access, framework, cloud administrators, users, owners, attackers, upload, manage files, authorize access, file audits, SQL database, integrity, availability, studies, reduce storage, large-scale data, robust security, data breaches..**

I. INTRODUCTION

In the era of explosive digital data growth, effectively managing information while safeguarding its security in cloud environments presents a formidable challenge. Addressing this critical need, TEXT SAFE introduces an innovative secure data deduplication system leveraging Advanced Encryption Standard (AES) for encryption and decryption processes. This framework aims to enhance both storage efficiency and data security by ensuring that textual data stored in the cloud remains confidential and shielded from unauthorized access.

Central to TEXT SAFE's architecture is its division of roles among cloud administrators, users, owners, and potential attackers, each with distinct responsibilities and permissions. Owners, for instance, have the capability to securely upload and manage files, authorize user access, and conduct comprehensive file audits. Meanwhile, users navigate stringent security protocols to request access to these files, bolstered by AES encryption protocols that safeguard data integrity.

Backing these functionalities is a robust SQL database infrastructure that supports seamless operations, maintaining data integrity and availability at all times. Studies underscore the significant potential of data deduplication, highlighting its capacity to reduce storage needs by an impressive 90-95% (Kwon et al., 2020; Hur et al., 2016). Such efficiency not only optimizes resource allocation but also fortifies defenses against data breaches, positioning TEXT SAFE as an indispensable solution for secure cloud storage environments in the modern digital landscape.

A. Problem Statement.

The exponential growth of textual data in cloud environments poses challenges in Maintaining both storage efficiency and security. Current systems often struggle to manage large-scale data while ensuring confidentiality and protection against unauthorized access. TEXT SAFE proposes a secure data deduplication framework utilizing Advanced Encryption Standard (AES) for encryption and decryption processes. This system aims to enhance storage efficiency by up to 95% while providing robust security measures to safeguard sensitive data. The framework delineates roles for cloud administrators, owners, users, and potential attackers, ensuring secure file management, access authorization, and audit capabilities. Effective implementation of TEXT SAFE addresses critical issues in cloud storage, offering a comprehensive solution for data integrity and security.

B. Existing System

In the existing system, data deduplication processes are implemented without leveragingadvanced encryption algorithms like AES, which results in potential vulnerabilities in datasecurity and confidentiality. To address this gap, our proposed system, DEDUP SHIELD, introduces a robust data deduplication framework utilizing AES for encryption and decryption processes. By integrating AES, DEDUP SHIELD ensures that textual data stored in the cloud remains confidential and protected from unauthorized access, significantly enhancing storage efficiency while mitigating risks associated with data breaches. This approach delineates specific roles for cloud administrators, users, owners, and potential attackers,

providing stringent security protocols for file access and management. Studies have shown that such AES-based systems can reduce storage needs by 90-95%, emphasizing the system's effectiveness in managing large-scale data securely (Kwon et al., 2020; Hur et al., 2016).

II. PROPOSED SYSTEM

A. Architecture of Proposed System.

The objective of the project "TEXT SAFE" is to develop a secure data deduplication system for cloud environments, utilizing Advanced Encryption Standard (AES) for encryption and decryption. This system aims to enhance storage efficiency by leveraging data deduplication, reducing storage needs by 90-95%. By ensuring confidentiality through AES encryption, TEXT SAFE aims to protect textual data from unauthorized access and mitigate risks of data breaches. Specific roles for cloud administrators, users, owners, and potential attackers are defined, enabling secure file management, user access authorization, and file audits. The project seeks to maintain data integrity and availability using an SQL database, offering robust security measures for large-scale data management in cloud storage.

B. Advantages of Proposed System.

- *Enhanced Data Security*
- *Improved Storage Efficiency*
- *Data Integration and Availability*
- *Mitigation of Security Risks*
- *Clear Role Definitions*

III. LITERATURE SURVEY

Secure data deduplication has been extensively studied as a critical component of efficient cloud computing. The investigate the use of natural language processing (NLP) techniques for deduplicating textual data. Their approach leverages semantic analysis to identify redundant information in unstructured datasets, significantly improving deduplication accuracy. By combining NLP with traditional deduplication methods, their solution is particularly effective for large datasets, such as those in research databases, news archives, and legal documents[1].

On the introduce with VeriDedup, a verifiable deduplication scheme designed to enhance trust in cloud service providers. Their approach ensures data integrity by integrating proof-of-duplication techniques, allowing users to verify the correctness of the deduplication process. This solution addresses concerns about potential malicious or erroneous operations by cloud providers and is particularly significant in high-transparency environments such as financial or healthcare systems [2].

The review on secure deduplication techniques and their applications, focusing on the balance between storage optimization and data security. They trace the evolution of deduplication methods from traditional chunking and hashing approaches to advanced encryption-aware schemes. The study underscores the growing need for cross-user deduplication frameworks that securely manage shared or identical data across users. Additionally, they explore the integration of deduplication with access control and authentication systems to enhance overall data security [3].

They have providea a comprehensive survey of secure deduplication methods, emphasizing the need to balance storage efficiency with robust security measures. Their work highlights the increasing demand for deduplication in multi-tenant cloud environments, where large volumes of data must be stored securely while minimizing redundancy. They identify encryption, hashing, and client-server deduplication architectures as essential components of modern solutions. Furthermore, they stress the importance of cross-user deduplication mechanisms that maintain data confidentiality even when identical data is stored by multiple users [4].

With propose of an innovative deduplication framework that combines data segmentation and advanced hashing techniques. By segmenting data into smaller chunks, their approach allows for more precise identification and removal of duplicates. This fine-grained segmentation not only enhances storage optimization but also ensures that encrypted data can undergo secure deduplication. Their solution leverages hashing algorithms to strengthen data integrity and demonstrates significant improvements in deduplication speed and network traffic reduction, making it highly suitable for cloud backup systems and archival storage [5] .

To address the challenges of managing encrypted cloud data with dynamic operations, such as updates and deletions. They propose a framework that integrates secure deduplication with dynamic data operations while maintaining consistent and reliable deduplication indices. Secure indexing mechanisms minimize performance overhead and ensure data integrity. This framework is particularly relevant for collaborative cloud applications and enterprise document management systems that require frequent data modifications[6].

To focus on the security challenges associated with cloud storage and deduplication, particularly in scenarios involving sensitive user data. Their study illustrates how deduplication can reduce storage costs and bandwidth usage but introduces vulnerabilities such as side-channel attacks and data leakage. To mitigate these risks, they recommend adopting encryption-aware deduplication schemes that secure data throughout its lifecycle. Additionally, they emphasize the importance of scalability to ensure deduplication solutions can handle dynamic and growing datasets effectively[7].

IV. CONCLUSION

Dedup Shield is a smart and secure way to manage cloud storage, making sure your data stays safe and organized. With the rapid growth of digital information, especially text-based files, it's easy to end up with multiple copies of the same thing, wasting space and increasing costs. That's where Dedup Shield steps in—it identifies and removes duplicate data while keeping everything encrypted with AES technology, so even if someone gains access, they can't read your files without the right key. It's like having a digital lockbox for your most important information. Looking ahead, Dedup Shield plans to introduce adaptive encryption, which automatically adjusts security levels based on how sensitive or frequently used your data is, and blockchain integration, adding a tamper-proof record of all data interactions for extra transparency and security. Whether you're a business looking to cut storage costs or an individual who values privacy, Dedup Shield offers a simple, secure, and efficient way to protect what matters most.

REFERENCES

[1] K. Ghassabi, P. Pahlevani, and D. E. Lucani, "Deduplication of textual data by NLP approaches," in *Proc. IEEE 97th Veh. Technol. Conf. (VTC-Spring)*, Florence, Italy, Jun. 2023, pp. 1–6, doi: 10.1109/vtc2023-spring57618.2023.10199538.

[2] X. Yu, H. Bai, Z. Yan, and R. Zhang, "VeriDedup: A Verifiable Cloud Data Deduplication Scheme With Integrity and Duplication Proof," 2022.

[3] M. A. Khan and M. A. Raza, "A Review on Secure Data Deduplication in Cloud Computing: Techniques and Applications," 2022.

[4] H. Patel and P. Thakkar, "Secure Data Deduplication in Cloud Computing: A Survey," 2021.

[5] S. Sharma and R. Kaul, "A Secure Deduplication Scheme for Cloud Storage Using Data Segmentation and Hashing," 2023.

[6] Y. Zhang, J. Liu, and X. Zhou, "Efficient Secure Deduplication for Encrypted Cloud Data with Dynamic Operations," 2022.

[7] P. Prajapati and P. Shah, "A review on secure data deduplication: Cloud storage security issue," *J. King Saud Univ. Compute. Inf. Sci.*, vol. 34, no. 7, pp. 3996–4007, Jul. 2022, doi: 10.1016/j.jksuci.2020.10.021.