Toward a Secure and Storage-Efficient Cloud Deduplication Framework for Multi-Tenant Environments

Dr S S Manikandasaran [1], Dr. SUDHAKAR K[2]

[1] Postdoctoral Researcher, Lincoln University College, Petaling Jaya, Selangor Darul Ehsan-47301, Malaysia; Email ID: pdf.Manikandasaran@lincoln.edu.my

[2] Professor / AI & DS, Nitte Meenakshi Institute of Technology, NITTE (Deemed to be University), Bengaluru, Karnataka, India.; Mail id: sudhakar.k@nmit.ac.in

Abstract: Cloud platforms increasingly store large volumes of repeated data generated through backup, synchronization, versioning, collaboration, and software-as-a-service workflows. In multi-tenant environments, duplicate content across users increases storage demand, input-output overhead, bandwidth consumption, and operational cost, while also raising challenges related to confidentiality, ownership assurance, metadata exposure, and tenant isolation. This paper addresses the need for a more integrated secure deduplication direction for multi-tenant cloud storage. A literature-grounded conceptual framework is proposed that combines adaptive deduplication, secure ownership verification, encrypted metadata management, and tier-aware storage optimization within a unified architecture. The paper synthesizes the prior work, identifies the major research gaps, and formulates the problem definition and research objectives that guide the proposed framework. The proposed framework is expected to support efficient storage reduction without weakening controlled access and multi-tenant security requirements. This work is relevant to enterprise cloud storage, backup systems, document repositories, and software-as-a-service platforms where repeated data and shared storage are common.

Keywords: Cloud deduplication; secure deduplication; multi-tenant cloud storage; proof of ownership; metadata protection

Introduction

Nowadays, cloud computing serves as a fundamental platform for digital data storage and management in substantial amounts. A large part of the data stored in the cloud is duplicated due to frequent uploading, which relates to data backup, synchronization, versioning, collaborative workflow, and software-as-a-service (SaaS). When speaking about multi-tenant clouds, this problem is enhanced because several users or organizations may store the same or similar data objects.

Data deduplication is acknowledged to be an important technique for eliminating storage redundancy. Using deduplication, it is possible to keep one physical instance of data duplicates and have multiple logical pointers to them, thus reducing costs of storage and bandwidth use [1]. The implementation of deduplication in the environment of multi-tenant clouds involves some issues. Duplicate detection provides knowledge about stored data. Data deduplication requires authentication of the ownership of data. Metadata becomes an overhead or the source of leakage. Tenant isolation should be ensured even when using the shared physical storage space. Thus, secure deduplication is not just an optimization problem but a general architecture and security problem [2].

The problem under investigation is the lack of a unified design that would take into account storage optimization, security, ownership confirmation, protection of metadata and application to a real multi-tenant cloud environment. The goal of this paper is to formulate the problem, review the literature,

identify major research gaps and develop a conceptual framework for the next stage of method designing and experiments.

Related work

Research works on cloud deduplication can be classified into five major streams, such as theoretical models of deduplication, encryption-compatible deduplication, proof of ownership methods, protecting metadata or indexing, and storage optimization. The work by Balaji et al. [3] is concerned with an approach for securing and deduplicating user data in public and private clouds. The importance of their research effort lies in drawing a clear relationship between data deduplication and the requirement for data security in different types of cloud deployments. Recently, much attention has been paid to the development of integrated security characteristics. In particular, Peng et al. [4] have proposed SecDedup with dynamic auditing and shown that secure deduplication is achievable in conjunction with a verifiable cloud-storage service.

Sabeerath et al. [5] have offered BTEDD – a technique based on block-level tokenization which is intended for providing efficient data deduplication in public clouds, thereby providing a valuable insight into the way block-level token generation contributes to improved storage efficiency. Song et al. [6] have developed a layered secure deduplication approach in cloud storage. Sabeerath et al. [7] have proposed ESPoW - an approach for efficiently verifying ownership and gaining secure access to deduplicated data in public clouds. All these works are relevant to the present paper, as all of them are centered around the issues of adaptive deduplication, proof of ownership and controlled access. However, they all consider particular approaches rather than develop an integrated multi-tenant framework.

The recent studies continue to show that secure deduplication in the clouds is shifting from duplicate removal to more advanced concepts of ownership verification, auditing, and dynamic data management. The research done by Dave et al. [8] provided an efficient method of ownership verification and its practical application in a cloud setting. Peng et al. [9] designed a blockchain-enabled solution for secure deduplication and cloud storage auditing, providing efficient fair arbitration and privacy. Lastly, Alfatmi and Sohani [10] designed a Hybrid SHA3 framework aimed at securing and optimizing client-side deduplication. Table 1 summarizes the main directions and limitations of representative prior work.

Table 1. Comparative summary of representative studies

Study	Main Focus	Key Contribution / Strength	Remaining Gap / Relevance to Present Work
Balaji et al. [3]	Data security and deduplication in public and private cloud	Proposed a framework for securing and deduplicating users' data across public and private cloud environments.	Relevant to the security and deduplication foundation of the present work, but does not specifically focus on adaptive multi-tenant deduplication with metadata and storage-tier optimization.
Peng et al. [4]	Secure deduplication with dynamic auditing	Combined secure deduplication with auditing support for cloud storage.	Integration with adaptive deduplication and multi-tenant storage placement is limited.
Sabeerath et al. [5]	Block-level token-based deduplication	Introduced BTEDD using block-level tokens to improve deduplication efficiency in public cloud infrastructures.	Supports the adaptive deduplication layer of the proposed framework, but broader integration with ownership verification, metadata protection, and auditing is limited.

Song et al. [6]	Layered secure deduplication	Proposed a layered secure deduplication model for cloud storage.	Adaptive decision-making and storage-tier optimization are not the central focus.
Sabeerath et al. [7]	Efficient and secured proof of ownership	Proposed ESPoW to enable authenticated access to deduplicated data in public cloud storage.	Highly relevant to the security and ownership layer of the present framework, but does not fully address multi-tenant storage optimization and secure metadata management.
Dave et al. [8]	Ownership verification in deduplicated cloud systems	Proposed a secure and efficient ownership verification scheme to ensure that a user must possess the complete file before proving ownership.	Highly relevant to the ownership layer of the proposed framework, but does not fully cover storage-tier optimization.
Peng et al. [9]	Blockchain-based secure deduplication and auditing	Proposed secure deduplication and cloud storage auditing with fair arbitration and privacy protection.	Useful for auditing and trust management, but blockchain-based schemes may introduce additional operational overhead.
Alfatmi and et al. [10]	SHA3-based secure client-side deduplication	Proposed a Hybrid SHA3-based framework for improving client-side deduplication security and efficiency.	Relevant to secure fingerprinting, but broader multi-tenant metadata management and storage placement remain open issues.

The reviewed studies show that secure cloud deduplication has progressed from basic duplicate elimination toward encryption-compatible deduplication, proof of ownership, access control, auditing, and dynamic ownership management. Recent works published in 2025 and 2026 further strengthen this direction by addressing ownership verification, data integrity auditing, blockchain-based arbitration, privacy protection, and secure client-side deduplication. However, the literature still shows a gap in developing an integrated framework that jointly considers adaptive deduplication, secure ownership validation, encrypted metadata management, tenant isolation, auditing, and tier-aware storage optimization in multi-tenant cloud environments. This gap motivates the conceptual framework proposed in the present paper.

Key Contributions

This paper makes four main contributions to secure and storage-efficient cloud deduplication in multi-tenant environments.

1. It defines the problem space of secure and storage-efficient deduplication in multi-tenant cloud environments by linking storage redundancy with confidentiality, ownership assurance, metadata protection, and tenant isolation.
2. It provides a focused literature synthesis that organizes prior research into major streams and highlights the limitations of fragmented approaches.
3. It identifies the major research gaps that motivate further work, namely the lack of integration across security, adaptability, multi-tenant deployment, and storage optimization.

4. It proposes a conceptual framework that unifies adaptive deduplication, secure ownership verification, encrypted metadata management, and tier-aware storage optimization as the basis for later prototype and validation stages.

Proposed Conceptual Framework and Research Method

This research work presents the proposed method and preliminary research design. The proposed framework is composed of six connected layers. Figure 1 illustrates the proposed conceptual framework, showing how tenant uploads are processed through adaptive deduplication, ownership assurance, secure metadata management, storage optimization, and controlled retrieval.

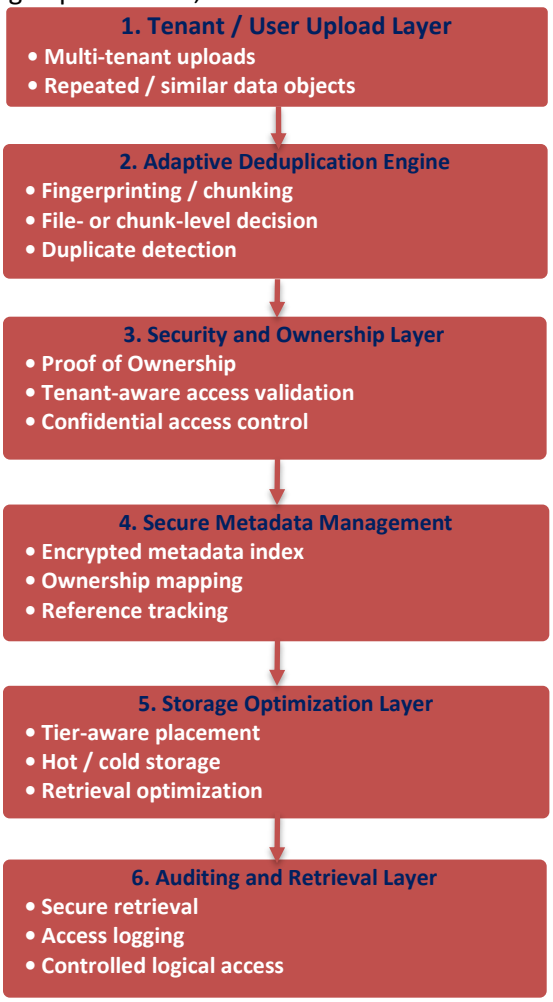

Figure 1. Proposed conceptual framework for secure and storage-efficient cloud deduplication in multi-tenant environments.

The first layer is the Tenant/User Upload Layer, where multiple tenants upload files or data objects that may contain repeated or similar content. The second layer is the Adaptive Deduplication Engine, which performs file fingerprinting or chunking and determines whether file-level or chunk-level deduplication is more suitable for the observed workload. The purpose of this layer is to avoid the limitations of static deduplication strategies.

Thirdly, the Security and Ownership Layer include proofs of ownership, tenant-aware access validation, and confidential access control. Its purpose is to prevent cases of false ownership and possible cross-tenant information disclosure. Fourthly, the Secure Metadata Management consists of encrypted

metadata indexing, ownership mapping, and reference tracking to help in controlled access and scalable implementation. Fifthly, the Storage Optimization Layer includes tier-aware storage allocation and decision of what to keep in hot/cold storage to increase practical efficiency by more than just duplicate elimination. Finally, the Auditing and Retrieval Layer allows for controlled logical access, secure retrieval and auditing of access.

The methodological approach for further stages will include the following five steps: requirement analysis, framework design, prototype development, experiments and evaluation and finally validation/refinement. The framework should be developed and tested in the cloud-based testbed environment and will be compared to the baseline solutions such as non-deduplicated storage, fixed file-level deduplication and fixed chunk-level deduplication. The evaluation will involve the assessment of the following parameters: storage savings, deduplication ratio, upload overhead, retrieval time, bandwidth savings, metadata overhead, and confidentiality.

Since this study is at the first conceptual stage, the main result achieved is the literature-based and problem-oriented architecture, which can be further developed and tested during the next stage.

Discussion

The literature review reveals that there is a very lively research direction devoted to secure cloud deduplication, while at the same time there is much room for improvement within the field. In earlier works, most attention was paid to deduplication or encryption compatibility, whereas proof of ownership was focused mostly on controlled access validation. Later studies introduced auditing, ownership management and layered secure deduplication, although their integration into the required context is still not sufficiently advanced.

This issue is especially important in the multi-tenant case, where a technique that increases deduplication efficiency but decreases data security is not suitable for use in cloud computing. In contrast, a solution that improves security but increases the number of metadata or negatively affects retrieval performance is unlikely to be used by cloud vendors. Thus, the issue is not whether there have been any studies in the field of secure deduplication, but whether they have been conducted with sufficient awareness of all dimensions involved.

The proposed theoretical framework takes into account the interrelation between deduplication, proof of ownership, metadata protection and storage optimization, which makes it applicable not only for further academic research, but also for practical implementation in enterprise cloud storage, backup solutions, digital repository systems and multi-tenant software-as-a-service. On the other hand, the proposed solution is purely conceptual in its current stage and cannot provide any performance benefits, which will be addressed in later research stages.

Conclusion

The current paper considers the problems of redundant storage of data in a multi-tenant cloud computing environment and the related security issues of confidentiality, ownership assurance, metadata exposure, and tenants' isolation. Based on the systematic analysis of existing research related to secure cloud deduplication, the paper develops a theoretical framework that incorporates the aspects of adaptive deduplication, secure ownership verification, encryption of metadata, optimized tiering of storage, and auditing and retrieval controls. The analysis reveals that the existing research provides some valuable insights in terms of encryption-compatible deduplication, proofs of ownership, access control and auditing; however, the existing solutions are quite incomplete and do not address the aspects of efficient storage in conjunction with the security requirements. Thus, the theoretical framework offered by the current paper serves as an initial point of departure for further research, where the architecture will be developed in prototype form and evaluated in comparison with baseline approaches based on such

performance parameters as storage savings, deduplication ratio, upload overhead, retrieval overhead, bandwidth reduction, and confidentiality assurance.

References

[1] W. Xia, H. Jiang, D. Feng, F. Douglis, P. Shilane, Y. Hua, M. Fu, Y. Zhang and Y. Zhou, "A comprehensive study of the past, present, and future of data deduplication," Proceedings of the IEEE, vol. 104, no. 9, pp. 1681-1710, 2016. doi: 10.1109/JPROC.2016.2571298.

[2] Y. Shin, D. Koo and J. Hur, "A survey of secure data deduplication schemes for cloud storage systems," ACM Computing Surveys, vol. 49, no. 4, Art. no. 74, 2017. doi: 10.1145/3017428.

[3] Balaji K and S. S. Manikandasaran, "Data Security and Deduplication Framework for Securing and Deduplicating Users' Data in Public and Private Cloud Environment," Journal of Scientific Research, vol. 14, no. 1, 2022, pp. 153-165. doi: 10.3329/jsr.v14i1.54063.

[4] L. Peng, Z. Yan, X. Liang and X. Yu, "SecDedup: Secure data deduplication with dynamic auditing in the cloud," Information Sciences, vol. 644, Art. no. 119279, 2023. doi: 10.1016/j.ins.2023.119279.

[5] Sabeerath, K., S. S. Manikandasaran, "BTEDD: Block-level tokens for efficient data deduplication in public cloud infrastructures," The Scientific Temper, vol. 15, no. 3, 2024, pp. 2507-2514. doi: 10.58414/scientifictemper.2024.15.3.16.

[6] M. Song, Z. Hua, Y. Zheng, H. Huang and X. Jia, "LSDedup: Layered secure deduplication for cloud storage," IEEE Transactions on Computers, vol. 73, no. 2, pp. 422-435, 2024. doi: 10.1109/TC.2023.3331953.

[7] Sabeerath, K., S. S. Manikandasaran, “ESPoW: Efficient and Secured Proof of Ownership Method to Enable Authentic Deduplicated Data Access in Public Cloud Storage,” The Scientific Temper, vol. 15, no. 4, 2024, pp. 3165-3172. doi: 10.58414/scientifictemper.2024.15.4.25.

[8] J. Dave, K. R. Ram, P. Patil, H. Patil, S. Borole and S. Panda, "Secure and efficient ownership verification for deduplicated cloud computing systems," *Journal of Cloud Computing*, vol. 14, Art. no. 20, 2025. doi: 10.1186/s13677-025-00743-y.

[9] X. Peng, W. Shen, Y. Gao *et al*., "Blockchain-based secure deduplication and cloud storage auditing with fair arbitration and privacy protection," *Cluster Computing*, vol. 28, Art. no. 541, 2025. doi: 10.1007/s10586-025-05196-1.

[10] K. Alfatmi and M. Sohani, "Hybrid SHA3 Junk Encryption based Crystal ZEUS framework for secure client-side deduplication on cloud," *Journal of Cloud Computing*, 2026. doi: 10.1186/s13677-026-00926-1.