Enabling Secure and Efficient Data Deduplication in Cloud Storage Systems Using Convergent Encryption and Bloom Filters

- Wei Chen, School of Computer Science and Engineering, South China University of Technology, Guangzhou, China

3. Abstract:
Data deduplication is a crucial technique employed in cloud storage systems to eliminate redundant data and optimize storage utilization. However, the process of identifying and removing duplicate data raises significant security and privacy concerns, as it may expose sensitive information to unauthorized parties. This research paper presents a novel approach to enable secure and efficient data deduplication in cloud storage systems using convergent encryption and Bloom filters. The proposed framework ensures data confidentiality while facilitating the identification and elimination of duplicate data across multiple users. The research methodology involves the development of a secure data deduplication protocol that combines convergent encryption and Bloom filters. Convergent encryption is employed to generate identical ciphertext for identical plaintext data, enabling the identification of duplicates without compromising data confidentiality. Bloom filters, on the other hand, are used to efficiently represent the set of encrypted data blocks and facilitate fast duplicate checks. The proposed approach is designed to handle both file-level and block-level deduplication, providing flexibility and granularity in eliminating redundant data. The framework is evaluated through extensive security analysis and performance measurements, demonstrating its effectiveness in preserving data confidentiality while achieving high deduplication ratios and minimizing storage overhead. The study presents a detailed comparison with existing data deduplication techniques, highlighting the advantages of the proposed approach in terms of security, efficiency, and scalability. The findings of this research have significant implications for cloud storage providers and users seeking to optimize storage utilization while maintaining data security and privacy. By enabling secure and efficient data deduplication, the proposed framework helps reduce storage costs, improve data management, and enhance the overall performance of cloud storage systems. This research contributes to the advancement of secure and privacy-preserving data deduplication techniques in the era of cloud computing, addressing the critical challenges of data confidentiality and storage efficiency in multi-tenant environments.

References

[1] K. Alwasel, Y. Li, P. P. Jayaraman, S. Garg, R. N. Calheiros, and R. Ranjan, “Programming SDN-native big data applications: Research gap analysis,” *IEEE Cloud Comput.*, vol. 4, no. 5, pp. 62–71, Sep. 2017.

[2] M. Abouelyazid, “Forecasting Resource Usage in Cloud Environments Using Temporal Convolutional Networks,” *Applied Research in Artificial Intelligence and Cloud Computing*, vol. 5, no. 1, pp. 179–194, Nov. 2022.

[3] M. Yousif, “Cloud-native applications—the journey continues,” *IEEE Cloud Comput.*, vol. 4, no. 5, pp. 4–5, Sep. 2017.

[4] M. Abouelyazid and C. Xiang, “Architectures for AI Integration in Next-Generation Cloud Infrastructure, Development, Security, and Management,” *International Journal of Information and Cybersecurity*, vol. 3, no. 1, pp. 1–19, Jan. 2019.