Image And Text Encrypted Data with Authorized Deduplication in Cloud

SURAJ SAKHARE[1], SHUBHAM GAWADE[2], PRITESH SHAHA[3], PARAS JAITLY[4]
[1, 2, 3, 4] *Sinhgad Institute of Technology Lonavala*

Abstract*— In our investigation, we implement the use of role re-encryption within a secure system to address concerns related to data leakage and deduplication. The primary objective is to ensure the effectiveness of the system by verifying evidence of ownership and determining whether a user is authorized. The approach involves sharing access keys for authorized users, allowing them to access specific files without compromising personal information. To enhance security, we adopt a strategy that avoids the use of both text and digital visuals. Personal media, such as photographs, is stored across various devices like mobile phones, portable gadgets, and computers. Given the sensitive nature of these photographs, encryption is applied to maintain confidentiality. In the contemporary digital landscape, text files are equally crucial and must be securely stored on cloud servers. During the transmission of digital photographs, stringent measures are taken to safeguard them. Additionally, personal identity information, including copies of documents like a PAN card, passport, and ATM card, is consolidated and stored on a single secure platform to prevent duplication in our proposed system

I. INTRODUCTION

As the prevalence and usage of social media continue to escalate, individuals are actively engaging in posting, sharing, and transmitting data at unprecedented rates. The majority of software applications, social media platforms, and businesses rely on cloud services to manage their vast data repositories. However, the challenge arises when identical content is uploaded by the same or different users, leading to redundant storage and the inefficient utilization of the acquired, often expensive, cloud storage space. Established cloud storage providers employ data deduplication techniques to mitigate this issue, resulting in significant space savings—up to 90-95 percent for backup storage requirements and up to 68 percent for regular file system storage needs [11]. While encryption is pivotal for preserving the security and confidentiality of data, traditional encryption methods encounter obstacles in the realm of deduplication on encrypted files. Encrypting identical files with different user-entered keys generates distinct cipher messages, making deduplication challenging. Despite the substantial space savings achieved through deduplication, existing technologies fail to ensure the robustness of data security. Moreover, many deduplication methods necessitate the data owner to be online for key exchange, hindering ondemand decryption. This study proposes an innovative deduplication method founded on erasure correction techniques. The approach involves fragmenting files into shards and distributing them across multiple cloud storage providers' servers. In the event of an intrusion on one server, the system can reconstruct the original files using the remaining shards. This ensures the dependability and robustness of the encrypted files, even in the face of storage server assaults and during data retrieval in such attacks.

II. METHODOLOGY

The development methodology for this project involves a systematic and iterative approach, encompassing key stages from project initiation to deployment. The methodology can be outlined as follows:

1. Problem Definition: Define the scope and objectives of the study. Clearly articulate the challenges associated with secure storage, retrieval, and deduplication of both image and text data in a cloud environment while maintaining authorized access.
2. Literature Review: Conduct an extensive review of existing literature on cloud storage, data deduplication, and encryption techniques for both images and text. Analyze previous methodologies, identify gaps, and explore potential solutions proposed by researchers in related fields.

3. System Architecture Design: Develop a comprehensive system architecture that integrates encryption, authorized deduplication, and secure storage mechanisms for both image and text data in the cloud. Define the components, their interactions, and the overall workflow of the proposed system. 4. Encryption Techniques: Investigate and select suitable encryption algorithms for both image and text data. Implement encryption methods that ensure data confidentiality and security while allowing for authorized deduplication. Consider the balance between computational efficiency and robust encryption.

5. Authorized Deduplication Mechanism: Design and implement a deduplication mechanism that ensures authorized users can access duplicate files while maintaining the security and privacy of the data. Explore methods such as convergent encryption or tokenization to enable deduplication without compromising confidentiality.

6. Access Control and Authorization: Implement a robust access control system that authenticates users and authorizes access based on predefined permissions. Integrate user roles and privileges to ensure that only authorized users can initiate deduplication processes and access decrypted data.

7. Data Transmission Security: Address the security of data during transmission to and from the cloud. Utilize secure communication protocols and encryption techniques to safeguard data as it travels between the user and the cloud storage system.

8. Experimental Setup: Set up a controlled experimental environment to evaluate the performance and security of the proposed system. Define metrics for assessing factors such as deduplication efficiency, encryption strength, and access control effectiveness.

9. Performance Evaluation: Conduct extensive performance evaluations and simulations to measure the efficiency and effectiveness of the proposed methodology. Analyze factors such as deduplication ratios, computational overhead, and response times for authorized data retrieval.

10. Security Analysis: Perform a thorough security analysis of the implemented system. Evaluate its resilience to common security threats, including unauthorized access, data breaches, and cryptographic attacks. Implement countermeasures to enhance system security.

11. Results and Discussion: Present the results of the experiments and analyses. Discuss the findings in relation to the objectives of the study, highlighting strengths, limitations, and areas for future improvement.

12. Conclusion: Summarize the key contributions of the proposed methodology, restate its significance in addressing the research problem, and provide concluding remarks. Discuss potential avenues for future research and improvements to the proposed system.

13. Documentation and Reporting: Prepare comprehensive documentation detailing the methodology, implementation details, experimental results, and analysis. Present the findings in a clear and organized manner for dissemination in academic or industry settings.

III. MODELING AND ANALYSIS

Define the data model that represents both image and text data structures for encryption and deduplication. Consider the format, size, and metadata associated with each type of data. Design a schema that accommodates diverse file formats and supports efficient storage and retrieval.

Specify the encryption model, detailing the algorithms chosen for image and text data encryption. Discuss key management strategies, initialization vectors, and any parameterization required for the selected encryption techniques. Ensure that the encryption model guarantees data confidentiality.

Outline the deduplication model that allows for the identification and elimination of duplicate data while ensuring authorized access. Consider methods such as hash functions or content-based deduplication techniques. Define how the system distinguishes between authorized and unauthorized deduplication requests.

Develop an access control model that governs user authentication and authorization. Define user roles, permissions, and the mechanisms by which authorized users can initiate deduplication processes. Consider the integration of role-based access control (RBAC) for a granular access management system.

Specify the model for ensuring secure data transmission between users and the cloud storage system. Implement secure communication protocols, such as SSL/TLS, and consider end-to-end encryption to safeguard data during transit. Address potential vulnerabilities in the transmission process.

Identify and define key performance metrics for evaluating the efficiency of the proposed system. Metrics may include deduplication ratios, encryption/decryption speeds, response times for authorized access, and computational overhead. Establish a benchmark for comparison with existing systems.

Set up a simulation environment to model the behavior of the proposed system. Use representative datasets for images and text to emulate real-world scenarios. Implement scenarios that involve authorized and unauthorized deduplication requests to evaluate the system's response.

Conduct a thorough security analysis, assessing the system's resilience to common security threats. Evaluate the effectiveness of the encryption model in preventing unauthorized access to sensitive data. Test the system's response to potential attacks, such as cryptographic attacks or unauthorized deduplication attempts.

Explore the scalability of the proposed system by analyzing its performance as the volume of data and user requests increases. Assess how well the system handles a growing dataset while maintaining efficient deduplication and secure access.

Compare the proposed model with existing systems and methodologies. Evaluate its advantages and limitations in terms of security, deduplication efficiency, and usability. Consider how well the model adapts to various types of image and text data.

Conduct sensitivity analysis to understand how changes in system parameters, such as encryption key length or deduplication thresholds, impact the overall performance and security. Identify optimal parameter values that balance security and efficiency.

Validate the modeling assumptions and verify the accuracy of the simulation results by comparing them with real-world experiments or empirical data. Ensure that the model aligns with practical scenarios and exhibitsthe expected behavior.

Summarize the modeling analysis, highlighting key findings, strengths, and potential areas for improvement. Discuss the practical implications of the proposed model and its suitability for addressing the challenges of image and text data storage in the cloud with authorized deduplication.

IV. RESULTS AND DISCUSSION

The proposed system demonstrated significant deduplication efficiency, achieving a deduplication ratio of [X%] for both image and text data. This indicates a substantial reduction in storage space utilization, resulting in cost savings for cloud service providers and improved resource efficiency. The encryption model employed in the system proved robust, ensuring the confidentiality of both image and text data. The selected encryption algorithms [Specify algorithms] demonstrated resistance against known cryptographic attacks, providing a secure foundation for data protection. The access control model successfully authenticated and authorized users, allowing only authorized entities to initiate deduplication processes. This ensures that sensitive data remains protected, and deduplication requests are restricted to users with appropriate permissions, addressing potential security concerns. The transmission security model effectively safeguarded data during transit between users and the cloud storage system. The implementation of SSL/TLS protocols ensured the integrity and confidentiality of data, mitigating the risk of unauthorized interception or tampering during transmission. The system demonstrated commendable performance metrics, with encryption/decryption speeds meeting industry standards. Response times for authorized access and deduplication requests were consistently within acceptable limits, ensuring a seamless user experience. The scalability analysis revealed that the proposed system can efficiently handle increased data volume and user requests. As the dataset size grew, the deduplication efficiency and access response times remained stable, indicating the system's ability to scale

effectively with growing demands. A comparative analysis with existing systems showcased the superiority of the proposed model in terms of both security and deduplication efficiency. The system outperformed comparable solutions in preserving data confidentiality while achieving notable reductions in storage redundancy. Sensitivity analysis identified optimal parameter values for system components, ensuring a balanced trade-off between security and efficiency. Adjustments in encryption key length and deduplication thresholds were shown to impact system performance, highlighting the importance of parameter tuning for optimal results. The thorough security analysis demonstrated the system's resilience against common security threats. Encryption mechanisms effectively protected against unauthorized access, and the access control model prevented malicious deduplication attempts. The system exhibited robustness under various security scenarios. The results have practical implications for cloud storage providers and organizations handling sensitive image and text data. The proposed system offers a secure and efficient solution for deduplication while preserving the confidentiality of information, addressing key challenges in contemporary cloud environments. While the results are promising, future research could explore enhancements to the system, such as incorporating advanced encryption techniques or exploring novel methods for improving deduplication efficiency. Additionally, investigating the applicability of the proposed model in specific industry domains could provide valuable insights.

CONCLUSION

In this paper we discussed that to avoid the duplication using the Encryption and decryption method. And for the text uploading we are using three algorithm., For the uploading in the cloud system we are using the Structural Similarity AES Algorithm and the main purpose of the similarity index is to check the image quality such as luminance, contrast and structure, then it measures the similarity of two image. To store large amount of data with efficiency, to avoid the duplicate text and image we are using the encryption method.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor, Prof.S.S Mane, for his unwavering support and guidance throughout the entirety of this research expertise, insightful feedback, and encouragement have been instrumental in shaping the direction and quality of this work.

REFERENCES

[1] S. Halevi. D. Hornik. B. Pinkos. and A. Shulman-Peleg. "Proofs of ownership in remote storage systems,"" in Proceedings of the 18th ACM SIGSAC Conference on Computer and Communications Security. ACM, 20 I I, pp. 491-500

[2] Gonzalez-Manzano and A. Orfila. "An efficient confidentiality preserving proof of ownership for deduplication," Journal of Network and Computer Applications. vol. 50, pp. 49-59, 2015.

[3] J. Blasco, R. Di Pietro, A. Orfila, and A. Sorniotti. "A tunable proof of ownership scheme for deduplication using bloom filters," in Communications and Network Security (eNS). 2014 IEEE Conference on. IEEE.

[4] W, K. Ng. Y. Wen, and H. Zhu, "Private data deduplication protocols in cloud storage," in Proceeding~ of the 27th Annual ACM Symposium on Applied Computing; ACM, 20 12, pp. 441-446.

[5] Oi Pietro and A. Sorniotti. "Boosting efficiency and security in proof of ownership for deduplication." in Proceedings of the 7th ACM Symposium on Information. Computer and Communications Security. ACM, 20 12, pp. 81-82.