Text Encryption With Authorized Deduplication In Cloud

Prof. Yogesh Shepal [1], Rushikesh Deshmukh [2], Himanshu Barhate [3], Pooja Daundkar [4]
Computer Engineering Department [1,2,3,4]
Nutan Maharashtra Institute of Engineering and Technology Pune, Maharashtra [1,2,3,4]

Abstract — To enhance storage efficiency in cloud environments, the AES encryption scheme has been introduced, leveraging a key derived from the message itself for encryption purposes. This approach ensures that identical plaintexts result in identical ciphertexts. The AES scheme is further refined to include simultaneous encryption and provides comprehensive security definitions. Additionally, the MD5 algorithm, a cryptographic hashing method, is employed for digital signatures, content verification, and message integrity assurance. This hash-based method ensures that both sender and receiver obtain an identical file during transmission. Cloud computing facilitates the sharing of vast data volumes across networks. Numerous techniques exist to secure data in the cloud, yet recent methods show improved performance in handling encrypted content. Consequently, we propose a system for collecting, sharing, and securely distributing data with multi-owner privacy preservation in cloud settings. In this system, data owners can securely transmit private information to a selected group of users via the cloud

Index Terms—MD-5 (Message-Digest Algorithm), AES Algorithm.

I. INTRODUCTION

The title "Text Encrypted Data with Authorized Deduplication in Cloud" articulates the primary goals of enhancing cloud security using cryptographic approaches. This paper elaborates on key components aimed at improving how security is managed in cloud environments.

The proposed system utilizes a distributed computing framework to leverage the vast storage capabilities of the cloud, allowing authorized users to access the system reliably from anywhere, provided they have strong internet connectivity. With the significant increase in multimedia content online, secure deduplication techniques are essential to conserve space on cloud servers.

Initially, we introduce an Advanced Encryption Scheme which derives its encryption key from the message itself. This ensures that identical plaintexts result in identical ciphertexts. Our enhanced AES includes elements of convergent encryption and provides comprehensive security definitions. Cloud computing represents a step forward in sharing extensive data volumes across the network. Current data security methods in the cloud surpass older encryption-based techniques. Our proposal focuses on enabling data collection, sharing, and controlled distribution with privacy considerations for multiple owners within the cloud. This allows data owners to securely share sensitive information with specific user groups via the cloud.

II. LITERATURE SURVEY

Cyber-Physical Systems (CPS) integrate cyber elements, such as mobile users, with cloud computing to facilitate real-time data transfer.[7] In such systems, cloud storage utilizes data deduplication techniques to optimize storage space and bandwidth. Data deduplication in this context works by removing redundant data, thereby enhancing the efficiency of CPS operations. However, this approach poses significant security and privacy risks. For instance, using distinct encryption keys for different users can interfere with effective data deduplication.[8] Despite extensive research in this field, existing solutions often struggle with securing data while maintaining high performance and connectivity. To address these issues, we introduce a novel protocol called Switch (Secure Homomorphic Encryption for Webs of Cloud-Hosted data). This protocol uniquely combines secure encryption and data deduplication within cloud storage environments. Switch distinguishes itself as the inaugural protocol that allows for resilient, encrypted deduplication using only two-party cryptographic interactions.[7] Our detailed numerical analysis demonstrates that Switch not only achieves higher performance but also improves practicality when compared to existing methods documented in scholarly literature.

The focus is on the re-encryption deduplication storage system and identifying a vulnerability in the recently developed lightweight rekeying-aware encrypted deduplication scheme (LRAED), which we term the "end-reserved attack."[11] To address this, we introduce a secure data deduplication strategy that incorporates effective encryption through the use of convergent all-or-nothing transforms (CAONT) and randomly selected bits from a Bloom filter. Leveraging the intrinsic characteristics of one-way hash functions, our proposed scheme effectively thwarts the end-reserved attack and ensures the confidentiality of data owners' sensitive information. Moreover, our approach minimizes computational burdens as data owners are required to re-encrypt only a small segment of

the data package using CAONT, instead of the entire file.[8] Comprehensive security analysis and empirical results validate that our method is secure.

The commonly utilized file-level deduplication to our newly proposed block-level deduplication within cloud data centers [1]. Applied both deduplication strategies to a specific dataset and found that our block-level deduplication method outperforms the file-level approach by 5%. Furthermore, we expect that the efficacy of our approach will increase when tested on larger datasets with more users operating in similar environments.[2]

A fog-computing-enhanced mobile crowdsensing framework that leverages fog nodes to optimally assign tasks based on user mobility, enhancing the accuracy of task distribution.[4] Additionally, we present a fog-supported secure data deduplication strategy (Fo-SDD) designed to boost communication efficiency while preserving data privacy.[6] We have developed a BLS-based pseudo-random function specifically for fog nodes, allowing them to identify and eliminate duplicated data in collected reports without accessing the actual content of those reports.[12]

To protect the privacy of mobile users, the Fo-SDD has been further enhanced to anonymize user identities during the data collection phase.[7] This is achieved using the Chameleon hash function, which facilitates secure token generation and cost reimbursement for users who remain anonymous.[1] Our results confirm that both implemented schemes provide secure and efficient data deduplication.

A secure data deduplication scheme with an effective PoW process for dynamic power operation. Especially, our scheme supports both cross-level deduplication, we construct a new PoW scheme to ensure the label thickness and achieve the collective power verification. also, we design a lazy update strategy to achieve effective power operation.[4] For inside-stoner block-position deduplication, the stoner-backed key is used to realize coincident crucial operations and reduce the crucial storehouse space. Eventually, the security and performance analysis demonstrate that our scheme can ensure data confidentiality and label thickness, and it's effective in data power operation.[12]

III. METHODS

Administrators are required to log in with a valid username and password to manage the system. Once logged in, they can access extensive administrative features such as viewing and managing user details (including names, email addresses, and permissions), overseeing all e-commerce site permissions, and reviewing all products along with their prices and user reviews. They can also view historical data on products, analyze search terms and results, and assess product search comparisons and analytics.

For users, registration is mandatory before accessing the system. After registering and logging in with their credentials, users can perform various activities such as adding products, viewing detailed product reviews, browsing through upcoming and historical products, and making purchases. Users can also manage their accounts, conduct product searches by content, and review their search activities.

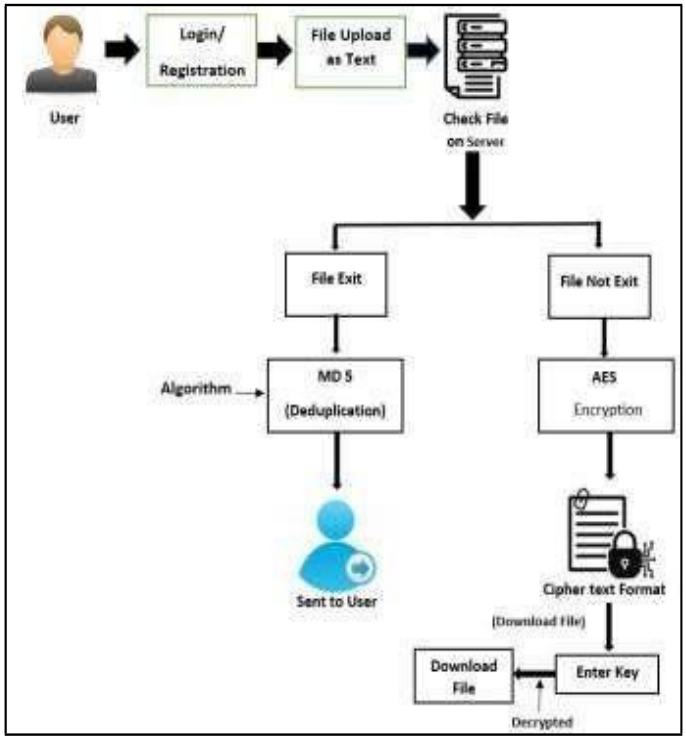

Fig. 1 System Architecture

The operation of the system is based on the fact that the text shown in the picture has some unique properties. Using the MD5algorithm. If deduplication occurs in the file, we resend it to the user; if the file does not contain deduplication, we keep the file.

IV. RELATED WORK

In the proposed system, authorized users can access the cloud platform through role-based key generation, ensuring that only they can retrieve data. Should an unauthorized user obtain a hash key, additional security measures are recommended to mitigate this risk.

This approach involves text deduplication at the storage level, with the client demonstrating to the server that it can store data without compromising its confidentiality. A cryptographic method is suggested to enhance cloud server security and manage deduplication more effectively. This method aims to filter out unnecessary files, thus boosting both cost efficiency and system performance.

Cloud Deduplication is a system that uses block-level deduplication integrated with convergent encryption to enhance security and optimize storage space in cloud environments. Similarly, Li et al. introduced the CD Store scheme to secure data deduplication, further contributing to efficient and secure cloud storage solutions

Pietro et al. introduced a Proof of Ownership scheme that utilizes random bits to verify file ownership by users. Blasco et al. focused on verifying authorized users to improve both the efficiency and security of cloud servers.

Gonzalez-Manzano highlighted the importance of convergent encryption (CE) in addressing data privacy and deduplication in cloud storage, a technique that encrypts data blocks based on ownership schemes.

V. ALGORITHM

AES (Advanced Encryption Standard) is a symmetric block cipher that encrypts data in fixed-size blocks of 128 bits using cryptographic keys of 128, 192, or 256 bits. This widely adopted protocol ensures data security by applying multiple rounds of transformations and substitutions, which vary based on the key length, to convert plaintext into ciphertext using the same key for both encryption and decryption processes.
Symmetric Key Algorithm: AES uses the same key for both encrypting and decrypting data, which simplifies key management yet requires secure key distribution.

Block Cipher: It processes data in fixed-size blocks (128 bits or 16 bytes). The input plaintext and output ciphertext are of equal length and divided into these block sizes.

Key Lengths: AES supports multiple key lengths - 128, 192, or 256 bits. Longer keys provide stronger security but may result in slower processing times.

Security: AES is considered highly secure. It has been analyzed extensively and is used by the U.S. government and other entities worldwide for classified information.

MD5 (Message Digest Algorithm 5) is a widely used cryptographic hash function that produces a 128-bit (16-byte) hash value from any input data, typically used for ensuring data integrity. Despite being found vulnerable to certain cryptographic attacks (making it unsuitable for secure cryptographic applications), MD5 is still useful in non-security-critical applications like checksums and identifying unique data sets – such as in deduplication processes.

Mathematical Model

Let S be the Whole system S= {I, P, O}
I-input
P-procedure
O-output

Input(I)
I= {Text Dataset}
Were,
Dataset-> text as .CSV, .TXT or any Format file
Procedure (P),
P= {I, Using I System perform operations and calculate the prediction}
Output(O)-
O= {System Store file securely and access securely.}

VI. CONCLUSION

The study of various methods to avoid the Deduplication using the Encryption and decryption method. For the text uploading we are using the two algorithms, For the uploading in the cloud system we are using the AES Algorithm. To store huge amounts of data efficiently, and to avoid duplicate text and images we are using this encryption technique.

REFERENCES

[1] Q. Li, J. Ma, R. Li, X. Liu, J. Xiong, and D. Chen, "Secure, efficient and revocable multi-authority access control system in cloud storage," Computers Security, vol. 59, pp. 45–59, 2016.

[2] J. Li, H. Yan, and Y. Zhang, "Certificate less public integrity checking of the group shared data on cloud storage," IEEE Transactions on Services Computing, pp. 1–12, 2018.

[3] J. Li, W. Yao, Y. Zhang, H. Qian, and J. Han, "Flexible and fine-grained attribute-based data storage in cloud computing," IEEE Transactions on Services Computing, vol. 10, no. 5, pp. 785–796, Sept 2017.

[4] J. Li, W. Yao, J. Han, Y. Zhang, and J. Shen, "User collusion avoidance cp-abe with efficient attribute revocation for cloud storage," IEEE Systems Journal, pp. 1–11, 2020.

[5] H. Wang, Z. Zheng, L. Wu, and P. Li, "New directly revocable attribute-based encryption scheme and its application in the cloud storage environment," Cluster Computing, vol. 20, no. 3, pp. 2385–2392, Sep 2020.

[6] C. Ma, Z. Yan, and C. W. Chen, "Scalable Access Control For Privacy-Aware Media Sharing," IEEE Transactions on Multimedia, vol. 21, no. 1, pp. 173– 183, Jan. 2019.

[7] K. Yang, Z. Liu, X. Jia, and X. S. Shen, "Time- Domain Attribute Based Access Control for Cloud- Based Video Content Sharing: A Cryptographic Approach," IEEE Transactions on Multimedia, vol. 18, no. 5, pp. 940–950, May 2016.

[8] H. Cui, R. H. Deng, Y. Li, and G. Wu, "Attribute-Based Storage Supporting Secure Deduplication of Encrypted Data in Cloud," IEEE Transactions on Big Data, pp. 1–1, 2019.

[9] M. J. Atallah, K. B. Frikken, and M. Blanton, "Dynamic and Efficient Key Management for Access Hierarchies," in Proceedings of the 12th ACM Conference on Computer and Communications Security, 2005, pp. 190–202.

[10] Q. Li, J. Ma, R. Li, X. Liu, J. Xiong, and D. Chen, "Secure, efficient and revocable multi-authority access control system in cloud storage," Computers Security, vol.59, pp.45–59,2016

[11]. Y. Zhang, X. Chen, l. Li. D. S. Wong, H. Li, and I. YOll, "Ensuring attribute privacy protection and fast decryption for outsourced data security in mobile cloud computing," Information Sciences, vol. 379. PI>. 42-61,2017.

[12] J. Xiong, J. Ren. L. Chen et al. "Enhancing privacy and availability for data clustering in intelligent electrical service of iot," IEEE Internet of Things Journal. vol. 6, no. 2, pp. 1530- 1540, Apri12019.

[13] Q Prof. Yogesh Shepal, Hemant C. Chavan, Gaurav B. Khatate, Yogesh P. More. "A System To Filter Unwanted Messages From OSN Users Walls," International Journal for Research in Engineering Application & Management (IJREAM) ISSN : 2454-9150 Vol-03, Issue 03, Apr 2017.