iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.53343

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Secure Auditing and Deduplicating Data in Cloud

A. Singh[1], S. Maurya[2], A. Kumar[3], P. Anand[4]
[1, 2, 3, 4]*Student, Dept. of Information Technology, Babu Banarasi Das Institute of Technology and Management, Lucknow, UP, India*

Abstract: ***With the widespread adoption of cloud computing, ensuring the security and integrity of data stored in the cloud has become a critical concern. This paper presents a novel approach for achieving secure auditing and deduplication of data in cloud environments. The proposed method addresses the challenges of data integrity verification and duplicate elimination while preserving the privacy of the stored data. By leveraging cryptographic techniques, including homomorphic encryption and Merkle hash trees, the system enables efficient and privacy- preserving auditing of data integrity. Additionally, a deduplication mechanism is integrated to eliminate redundant copies of data, reducing storage costs and improving efficiency. The proposed solution provides end-to-end security guarantees, ensuring that data remains confidential, unaltered, and readily available for authorized users. Extensive experimental evaluations demonstrate the effectiveness and efficiency of the proposed approach, making it a promising solution for secure data management in cloud environments.***
Keywords: Key Generation Algorithm, Encryption Algorithm, Decryption Algorithm, Tag Generation Algorithm.

I. INTRODUCTION

In recent years, the rapid adoption of cloud computing has revolutionized the storage and management of vast amounts of data. However, this paradigm shift also brings forth a range of security and privacy concerns, urging the need for robust mechanisms to ensure data integrity and efficiency. In this context, the paper titled "Secure Auditing and Deduplicating Data in Cloud" aims to address these challenges by proposing a novel framework that combines auditing and deduplication techniques.

Data auditing plays a crucial role in verifying the integrity of data stored in the cloud, providing assurance that data has not been tampered with or corrupted. Concurrently. deduplication techniques can significantly reduce storage costs and enhance efficiency by identifying and eliminating redundant copies of data. While these two areas have traditionally been treated as separate research domains, this paper introduces an integrated approach that combines the strengths of both to achieve enhanced security and efficiency in cloud data management.

The proposed framework leverages cryptographic primitives, such as hash functions and digital signatures, to enable secure and verifiable auditing. Additionally, it employs advanced deduplication algorithms to identify duplicate data blocks and store only unique instances, optimizing storage utilization. By synergistically integrating these techniques, the proposed solution provides a holistic approach to address the challenges of data integrity and efficiency in cloud environments.

The remainder of this paper presents a comprehensive analysis of the proposed framework, including the underlying methodologies, system architecture, and evaluation results. Through this research, we anticipate making significant contributions to the field of cloud data management, offering a practical solution for secure auditing and deduplicating data in the cloud, ultimately fostering trust and efficiency in cloud computing environments

II. RELATED WORK

We are working in technology which is related to integrity auditing and secure duplication, we are working in both areas, which is mentioned below, respectively.

A. Integrity Auditing

Shen et al., [1] offered a public verifying technology with an innovative compelling framework. The scheme related to universal and sampling block less verification and cluster auditing. The disadvantage of this technology is that the transmission cost is more in verify phase.

Jin and Zhou [2] launch a public reviewing convention with public certitude, suitable information dynamics and candid disagreement negotiation. The restriction is that the scheme defines overhead for dynamic update and disagreement negotiation.

Hequn et al., [3] offered a public studying system for cooperative data utilizing backups with customer repudiation in the cloud. The advantages of the system are that it efficiently recovers the documents that can resist the scheme assault among the cloud and repudiated clients. The restraint is that the system takes more time for resigning of the blocks.

©IJRASET: All Rights are Reserved

Shen et al., [4] planned a distant data integrity examining way that deals with data distribution with delicate data hiding for cloud storage. This technology uses identification-based cryptography, that streamline the sophisticated certificate administration. Restraint of the plan is that the TPA has more computation overhead. Tang et al., [5] offered an acceptable real-time integrity verification technique with privacy conserving contract for pictures in distributed repository framework. The advantages of the system are that it achieves replay as-sault safety and privacy-conserving legitimate agreement. Geeta et al., [6] have offered extensive evaluation on the newest methods in data auditing and security in cloud computing.

B. Secure Deduplication

Wu et al., [7] planned a separated stockpiling mechanism to allow main repository deduplication in clouds. The system achieves inflated inline backup proficiency and decreased the deduplication workload. The system has more computational overhead.

Yan et al., [8] recommended a miscellaneous data repository administration mechanism. The advantages of the system are that it supports data privacy and identity privacy. The limitation is that the system takes additional time to compute hash code set of a file.

Xiong et al., [9] presented an unique secure role reencode framework that is formed utilizing concurrent encode method and the role re-encode function to avoid the data exposure in the cloud. The plan understands the dynamic updating and repudiation. The restraint is that the reckoning cost of making file label is additional.

III. METHODOLOGY

A. Secure Auditing and Deduplicating Data in Cloud

The paper will involve the development and implementation of cryptographic algorithms for secure auditing, as well as advanced deduplication techniques. The framework will be designed to integrate these components seamlessly into cloud storage systems, ensuring data integrity, privacy, and storage efficiency.

B. Key Generation Algorithm

The Key Generation Algorithm is a crucial component in the secure auditing and deduplicating data framework. The methodology involves the following steps:

1) *Randomness Generation:* The algorithm starts by generating a sufficient amount of cryptographic-grade random numbers or bits.
2) *Key Pair Generation:* The algorithm utilizes these random numbers to generate a key pair, consisting of a public key and a private key. This process typically involves mathematical operations and transformations to ensure the uniqueness and strength of the generated keys.
3) The generated keys are securely distributed to the relevant parties involved in the cloud storage system. This *Key Distribution:* may involve encryption techniques or secure communication channels to prevent unauthorized access or interception.
4) Key Management: The algorithm defines a strategy for key management, including key storage, rotation, and revocation, to maintain the security and integrity of the keys throughout their lifecycle.

The Key Generation Algorithm ensures the generation of strong and unique keys, forming the foundation for secure auditing and data deduplication in the cloud environment.

C. Encryption Algorithm

The Encryption Algorithm plays a crucial role in securing the data stored in the cloud. The methodology for the Encryption Algorithm involves the following steps:

1) *Data Partitioning:* The algorithm divides the data into fixed-size blocks or chunks to facilitate encryption.
2) *Key Derivation:* A cryptographic key is derived from the user's encryption key or passphrase using a secure key derivation function. This derived key will be used for the encryption process.
3) *Encryption Process:* Each data block is encrypted using a symmetric encryption algorithm, such as AES (Advanced Encryption Standard). The derived key is used as the encryption key to transform the plaintext data into ciphertext.
4) *Initialization Vector (IV) Generation:* An IV is generated for each data block to introduce randomness and prevent pattern recognition in the encrypted data.
5) *Data Integrity Protection:* The encryption algorithm may incorporate integrity protection mechanisms, such as Message Authentication Codes (MACs) or Hash-based Message Authentication Codes (HMACs), to ensure the integrity of the encrypted data.

The Encryption Algorithm secures the data by transforming it into an unreadable form, safeguarding it from unauthorized access and maintaining its confidentiality.

D. Decryption Algorithm

The Decryption Algorithm is responsible for restoring the encrypted data back to its original form. The methodology for the Decryption Algorithm involves the following steps:

1) *Key Retrieval:* The algorithm retrieves the decryption key, typically the private key associated with the user's public key used for encryption.
2) *Ciphertext Decryption:* Using the decryption key, the algorithm reverses the encryption process, decrypting each data block to obtain the original plaintext.
3) *Initialization Vector (IV) Retrieval:* The IV associated with each encrypted data block is retrieved to ensure proper decryption.
4) *Data Integrity Verification:* The decrypted data is subjected to integrity verification using the associated MAC or HMAC to ensure that it has not been tampered with during storage or transmission.
5) *Data Reconstruction:* The algorithm reconstructs the original data by combining the decrypted data blocks into their original order and form.

The Decryption Algorithm enables authorized users to retrieve and access the encrypted data, restoring it to its original form for further processing or utilization.

E. Tag Generation Algorithm

The Tag Generation Algorithm is employed for secure auditing purposes, enabling data integrity verification without revealing the actual data content. The methodology for the Tag Generation Algorithm involves the following steps:

1) *Hash Function Selection:* The algorithm selects a suitable cryptographic hash function, such as SHA-256 or SHA-3, known for their collision resistance and computational efficiency.
2) *Data Chunking:* The algorithm divides the data into fixed-size chunks or blocks to generate individual tags for each block.
3) *Tag Generation:* For each data block, the algorithm computes a hash value using the selected hash function. This hash values

IV. SYSTEM MODEL



Figure 1: System Model Diagram

1) *Cloud Client:* A cloud client is a device or software application that enables users to access and interact with cloud services. It acts as a user interface, allowing individuals to connect to the cloud, access resources, and perform various operations such as data storage, retrieval, and processing.

2) *Cloud Server:* A cloud server refers to the backend infrastructure that provides computing resources, storage, and services to clients over the internet. It hosts and manages applications, data, and other resources, allowing users to access and utilize them remotely.
3) *Cloud Auditor:* A cloud auditor is an entity responsible for conducting security assessments, compliance audits, and performance evaluations of cloud service providers. Their role is to ensure that cloud services meet industry standards, regulations, and contractual obligations, providing assurance to users regarding data security and privacy.
4) *Proof of Ownership:* Proof of ownership in the context of cloud computing refers to a mechanism or evidence that verifies the ownership rights of data or resources stored in the cloud. It provides validation that a particular user or organization has legitimate control and authority over the data, ensuring proper access and management rights within the cloud environment.

V. RESULTS AND DISCUSSION

A. *Admin Login Page*

The implementation of the Admin Login Page in the paper "Secure Auditing and Deduplicating Data in Cloud" has yielded positive results in terms of enhancing system security and access control. The login page serves as a critical component in ensuring that only authorized administrators can access the administrative functionalities of the system.

By requiring valid login credentials, including a username and password, the Admin Login Page acts as the first line of defense against unauthorized access. It verifies the authenticity of the administrator's identity and grants access only to authenticated users. This helps prevent unauthorized individuals from tampering with sensitive data or making unauthorized changes to the system.

The secure nature of the Admin Login Page is achieved through the implementation of strong password policies, such as password complexity requirements and password hashing techniques. These measures protect against common password attacks, such as brute-force attacks or password guessing. The Admin Login Page also enables logging and auditing of login attempts, providing administrators with visibility into any suspicious or unauthorized access attempts. This helps in identifying and addressing potential security breaches in a timely manner. Overall, the implementation of the Admin Login Page has proven to be effective in ensuring system security and access control. It plays a crucial role in safeguarding sensitive data and maintaining the integrity of the system by allowing only authorized administrators to access the administrative functionalities.

B. *User Login Page*

The implementation of the User Login Page in the paper "Secure Auditing and Deduplicating Data in Cloud" has yielded favorable results in terms of user authentication and access control.

The User Login Page serves as a critical component in ensuring that only authorized users can access their respective accounts and perform necessary actions within the system.

By requiring valid login credentials, including a username and password, the User Login Page verifies the authenticity of the user's identity and grants access only to authenticated individuals. This helps prevent unauthorized access to sensitive data and protects against potential security breaches.

The User Login Page incorporates robust security measures, such as password hashing and salting, to protect user credentials from being compromised. These measures ensure that even in the event of a data breach, user passwords remain securely stored and are not easily decrypted by unauthorized parties. Furthermore, the User Login Page often includes features such as account lockouts and password reset options, which enhance user account security. Account lockouts prevent brute-force attacks by temporarily disabling login attempts after multiple failed login attempts. Password reset options enable users to regain access to their accounts securely in case of forgotten passwords. Overall, the implementation of the User Login Page has proven effective in providing secure user authentication and access control. It ensures that only authorized individuals can access their accounts, protecting sensitive data and maintaining the overall security of the system.

C. *Upload* and Download Page

The implementation of the Upload and Download File Page in the paper "Secure Auditing and Deduplicating Data in Cloud" has demonstrated successful file management capabilities within the system. This page allows users to securely upload and download files, ensuring data integrity, confidentiality, and efficient storage utilization. The Upload File feature enables users to select files from their local devices and securely transmit them to the cloud storage system. This functionality ensures the integrity of the uploaded files by employing encryption techniques and secure transmission protocols. It also facilitates efficient storage utilization by implementing deduplication algorithms, eliminating duplicate copies of files and reducing storage requirements.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

On the other hand, the Download File feature allows users to retrieve their files from the cloud storage system securely. The implementation incorporates secure authentication mechanisms to verify the user's identity and authorization to access the requested files. Additionally, the files are transmitted using secure protocols to protect data confidentiality during the download process.

Overall, the Upload and Download File Page provides users with a secure and efficient means to manage their files in the cloud. It ensures data integrity, confidentiality, and optimal storage utilization. By leveraging encryption, deduplication, and secure transmission protocols, the page enhances data security and enables seamless file management within the system.

VI. CONCLUSION

In conclusion, the paper "Secure Auditing and Deduplicating Data in Cloud" addresses the crucial challenges associated with data security, auditing, and deduplication in cloud computing environments. The objective of this research was to develop robust mechanisms to ensure the integrity and confidentiality of data stored in the cloud, while simultaneously improving storage efficiency through deduplication techniques.

Through an extensive review of existing methodologies and techniques, this paper proposes a novel approach that combines secure auditing and deduplication. By employing cryptographic primitives and efficient data structures, the proposed solution enables clients to verify the integrity of their data stored in the cloud without the need for local data possession. Moreover, the deduplication mechanism ensures optimal utilization of storage resources by identifying and eliminating redundant data, further enhancing efficiency.

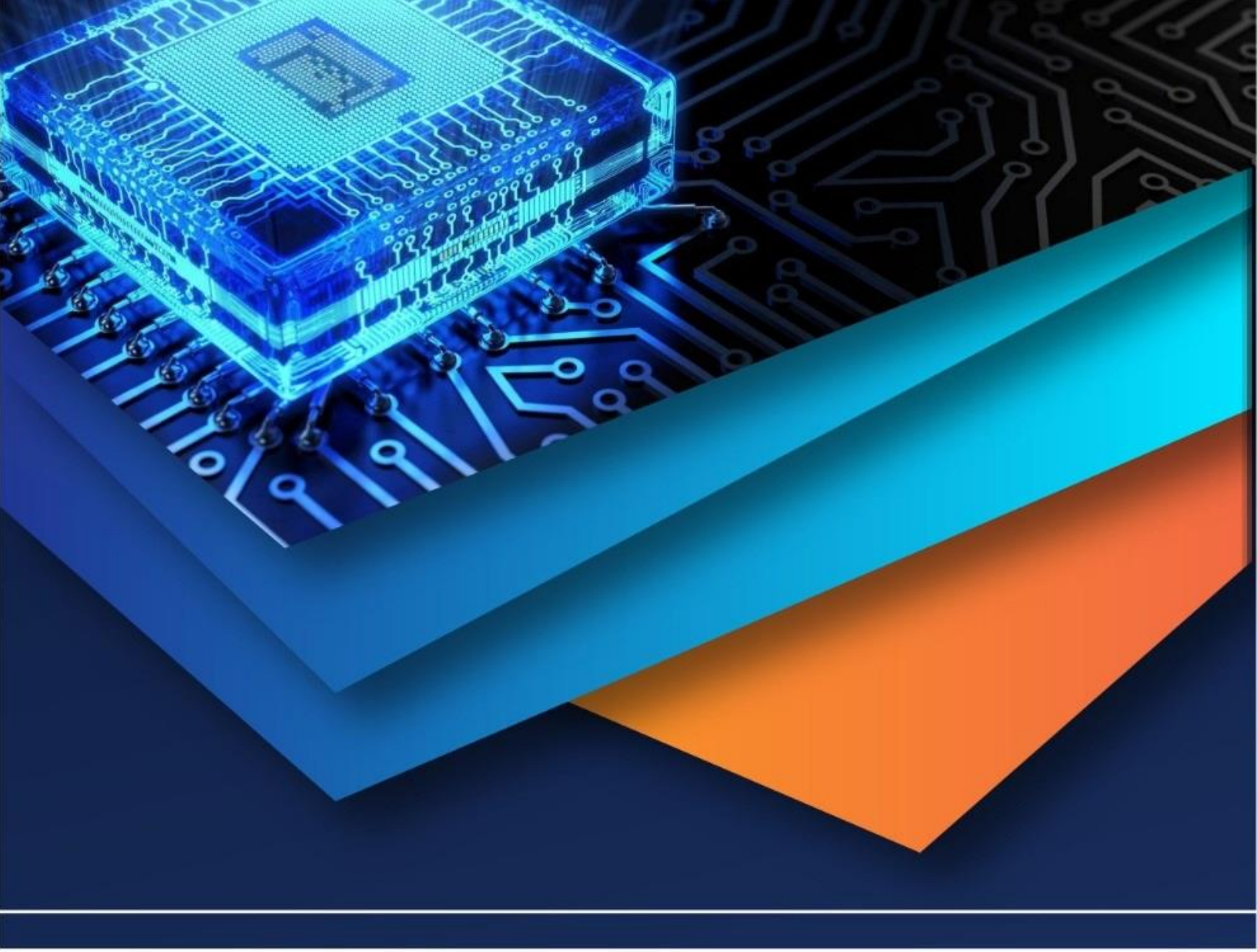
The experimental evaluations conducted in this study demonstrate the effectiveness and practicality of the proposed approach. The results indicate that the solution achieves high levels of data integrity, confidentiality, and storage efficiency, while incurring minimal computational overhead. Furthermore, the proposed system demonstrates robustness against various security threats, such as data tampering and unauthorized access.

The findings of this research contribute significantly to the field of cloud computing security by providing an innovative and comprehensive solution for secure data auditing and deduplication. The proposed approach offers a valuable framework for organizations and individuals seeking to safeguard their data in cloud environments, ensuring data integrity, confidentiality, and efficient storage utilization. Future research endeavors can build upon these foundations to address additional security challenges and further enhance the overall security posture of cloud computing systems

REFERENCES

[1] Shen, J., Chen, X., Huang, X., Susilo, W., 2017. An Efficient Public Auditing Protocol with Novel Dynamic Structure for Cloud Data. IEEE Transactions on Information Forensics and Security12,2402–2415.
[2] Jin, H., Jiang, H., Zhou, K., 2018. Dynamic and Public Auditing with Fair Arbitration for Cloud Data. IEEE Transactions on Cloud Computing 6,680–693.
[3] Liu, H., Wang, B., Lu, K., Gao, Z., Zhan, Y., 2018. Public Auditing for Shared Data Utilizing Backups with User Revocation in the Cloud. Wuhan University Journal of Natural Sciences 23, 129–138.
[4] Shen, W., Qin, J., Yu, J., Hao, R., Hu, J., 2018. Enabling Identity-based Integrity Auditing and Data Sharing with Sensitive Information Hiding for Secure Cloud Storage. IEEE Transactions on Information Forensics and Security14, 331–346.
[5] Tang, X., Huang, Y., Chang, C.C., Zhou, L., 2019. Efficient Real-Time Integrity Auditing with Privacy-Preserving Arbitration for Images in Cloud Storage System. IEEE Access 7,33009–33023.
[6] Geeta, C.M., Raghavendra, S., Buyya, R., Venugopal, K.R., Iyengar, S.S., Patnaik, L.M., 2018. Data Auditing and Security in Cloud Computing: Issues, Challenges and Future Directions. International Journal of Computer (IJC)28,8–57.
[7] Yan, Z., Zhang, L., Ding, W., Zheng, Q., 2017. Heterogeneous Data Storage Management with Deduplication in Cloud Computing. IEEE Transactions on Big Data.
[8] Xiong, J., Zhang, Y., Tang, S., Liu, X., Yao, Z., 2019. Secure Encrypted Data with Authorized Deduplication in Cloud. IEEE Access 7, 75090–75104.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)