iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** VI **Month of publication:** June 2023

DOI: https://doi.org/10.22214/ijraset.2023.53650

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 VI Jun 2023- Available at www.ijraset.com

Secure Deduplication with User-Defined Access Control in Cloud Storage

Vivek Solanke[1], Tushar Tambe[2], Swapnil Melkunde[3], Prof. Dr. S. R. Litake[4]
[1, 2, 3]*BE (E&TC) final year 2023, Electronics and Telecommunication, Department of Electronics andTelecommunication, PVGCOET and GKP(W)IoM, Pune-411009, Maharashtra, India*
[4]*Professor Department of Electronics and Telecommunication PVGCOET and GKP(W)IoM.*

***Abstract:** Cloud storage is one of the most important services of cloud computing, offering significant benefits to users by allowing them to outsource their data for storage and sharing with authorized parties. Secure deduplication has been extensively researched in cloud storage, aiming to eliminate redundancy in encrypted data and reduce storage space. Our proposed solution focuses on uploading files to the cloud storage system. When a file is uploaded, it is checked to see if an identical file already exists. If a duplicate is found, the system displays an entry indicating that the file already exists. This secure and efficient data deduplication scheme ensures that users can store their data securely in the cloud, with access limited to authorized parties. By eliminating redundancy, more space is available for uploading different files. To ensure data security, we utilize the MD5 algorithm for hashing purposes and AES for encryption. Our system offers a practical solution for organizations seeking to securely store their data in the cloud, while also preserving user privacy and data ownership.*

I. INTRODUCTION

Cloud storage has become an integral part of modern computing infrastructure, offering convenience, scalability, and cost-effectiveness. A key technique employed in cloud storage is data deduplication, which optimizes storage space and reduces data management costs. However, the use of deduplication raises concerns regarding data privacy and security, particularly when multiple users' data is stored in the same physical location. To address these concerns, various techniques, including encryption and access control, have been proposed. However, many existing solutions lack user-defined access control, which is crucial for safeguarding data privacy and security in multi-user environments. Data deduplication involves removing redundant data by storing only a single copy of each unique data segment. This approach effectively reduces storage costs and enhances the overall efficiency of cloud storage systems. Nevertheless, traditional deduplication methods often fall short in terms of security, as they fail to provide adequate protection for sensitive data. To tackle this challenge, our proposed system introduces a secure deduplication technique that combines encryption and user-defined access control to ensure the confidentiality and privacy of data. By integrating cryptographic techniques such as the MD5 algorithm and hash functions, our scheme achieves secure deduplication while maintaining data integrity. Access control, a critical component of cloud computing, has found widespread use in practical cloud products. However, existing schemes that address authorized secure deduplication by introducing an additional authorized server have been susceptible to duplicate faking attacks. To overcome this vulnerability, we present an efficient and secure cross-user deduplication scheme that incorporates user-defined access control, ensuring data confidentiality, tag consistency, access control, and resistance to duplicate faking attacks simultaneously. In our research, we propose a novel secure deduplication scheme with user-defined access control for cloud storage. Our scheme leverages a combination of cryptographic techniques, including the MD5 Algorithm and hash functions, to achieve secure deduplication and protect data privacy. By addressing the limitations of existing solutions, our approach provides an innovative solution for organizations seeking to ensure the confidentiality and security of their data in the cloud..

II. EASE OF USE

When it comes to ease of use for a Secure Deduplication with User-Defined Access Control in Cloud Storage system,there are a few key considerations to keep in mind:

1) *User Interface:* A well-designed user interface can make a complex system much easier to use. Consider using a simple, intuitive interface that allows users to easily navigate and interact with thesystem.
2) *Clear Instructions:* Provide clear instructions on how to use the system, including how to set up access control and how to securely store and retrieve data. Make sure that these instructions are easily accessible and written in plain language that is easy for users to understand.

3) *Automate Where Possible:* Automating certain taskscan help to reduce the burden on users and make the system easier to use. For example, you could automate the process of deduplicating files or managing access control permissions.
4) *Provide Support:* Finally, it is important to provide support to users who may have questions or issues with the system. This could include providing a helpdesk or support forum where users can askquestions and get help when they need it.

Overall, by focusing on user interface design, clear instructions, automation, existing tools, and providing support, you can make your Secure Deduplication with User-Defined Access Control in Cloud Storage project much easier for users to use.

III. LITERATURE SURVEY

In a recent study [1], an innovative and efficient secure deduplication scheme with user-defined access control is proposed. Unlike previous approaches, this scheme eliminates the need for an additional authorized server or hybrid cloud architecture, making it more streamlined and practical. The Content Security Policy (CSP) is leveraged to manage access rights without compromising data confidentiality. Moreover, the scheme incorporates the use of Bloom filters for efficient duplicate checks.

Thorough security analyses of the proposed scheme demonstrate its ability to achieve multiple security objectives simultaneously. It ensures data confidentiality, access control, tag consistency, and resistance against brute-force attacks.

In a separate paper [2], the focus is on Attribute-Based Encryption (ABE) as a suitable solution for fine-grained cryptographic access control. The challenge of user revocation is addressed through the implementation of a cipher text policy attribute-based scheme, specifically designed for cloud-enabled user revocation. This scheme provides comparable granularity to ABE while minimizing computation and communication overhead at the user's end. Another paper [3] proposes a scheme based on attribute-based encryption (ABE) for secure data deduplication and data access control in the cloud. The scheme supports digital rights management based on the data owner's expectations, saving storage space by storing only one copy of duplicate data. The scheme is scalable, enabling support for multiple duplication instances and large volumes of duplicated data. Identity-based cryptography is the topic of discussion in a comprehensive paper [4]. This cryptographic technique, related to public key cryptography, is explored for its feasibility and potential benefits in current and future environments. The paper highlights the advantages and limitations of identity-based cryptography and discusses its role in secure communication. It distinguishes between symmetric key cryptography, where a single key is used for encryption and decryption, and asymmetric key cryptography, where a public-private key pair is employed. Lastly, an enhanced MD5 algorithm with dynamic variable length and high efficiency is proposed in [5]. This method aims to simulate the highest level of security by generating different fragment sizes to thwart hostile attacks. The paper emphasizes the significance of the improved MD5 algorithm in maintaining data integrity, reliability, and authenticity. Various modifications, including the use of key technology, have been implemented to enhance the algorithm's collision resistance and resilience against penetration attempts.

IV. METHODOLOGY

A. *Explanation of Flow diagram*

1) *User:* A user refers to an individual who intends to utilize our system for secure deduplication with user-defined access control in the cloud. The user interacts with the system to manage their data and define access control policies.
2) *Encryption:* Encryption is the process of converting information or data into a coded form that can only be understood by someone who possesses the corresponding decryption key. It ensures the confidentiality and security of data by making it unreadable to unauthorized parties.
3) *Decryption:* Decryption is the reverse process of encryption, where the encrypted data is transformed back into its original, readable form using the appropriate decryption key. It allows authorized users to access and understand the encrypted data.
4) *AES (Advanced Encryption Standard):* AES is an encryption algorithm widely used in our system. It provides a strong level of security and is commonly employed to protect sensitive data. AES operates on fixed-size blocks of data and employs symmetric key encryption, where the same key is used for both encryption and decryption.
5) *MD5 (Message Digest 5):* MD5 is a cryptographic hash function utilized in our system. It generates a fixed-sized (128 bits) output, known as a hash value or digest, for input data of any length. MD5 is commonly used to verify data integrity and as a checksum in various applications.
6) *Hash Function:* A hash function is a mathematical function that takes an input (data of any size) and produces a fixed-size output, called a hash value or digest. The primary characteristics of a hash function include generating a unique hash value for each unique input, being deterministic (same input produces the same hash), and being computationally efficient. Hash functions are widely used in cryptography, data integrity checks, and digital signatures.

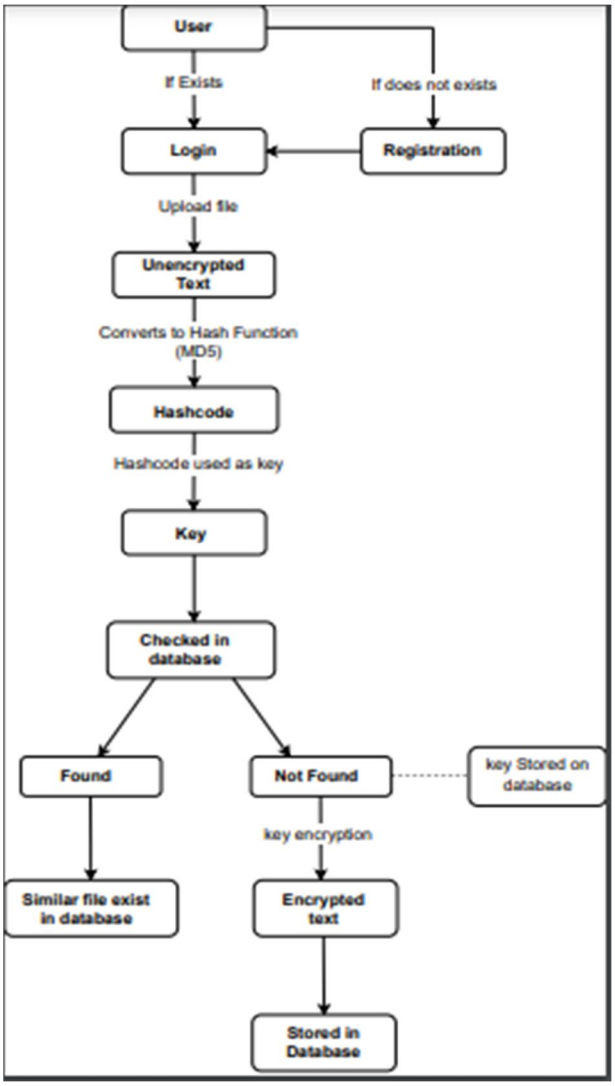


Fig.1 Flow Diagram

B. Working

1) User need to register to the system.
2) After registration, login to the system.
3) After successful login, user should upload his/her file.
4) Uploaded file will be converted to unique hash code by MD5 hash function.
5) Hash code as a key will be searched on database of the system.
6) If same key is not found then, uploaded file will be encrypted and stored on the database.

If same key is found, meaning similar file already exist in the system. Therefore that file will not be stored in the system.
User can download file from the system any time.

V. ALGORITHMS

A. AES (Advanced Encryption Standard)

The Advanced Encryption Standard (AES) is a widely-used encryption algorithm that can be applied in the Secure Deduplication with User-Defined Access Control in Cloud Storage project to ensure secure storage of data. AES is a symmetric encryption algorithm, meaning that the same key is used for both encryption and decryption. It offers a high level of security, confidentiality, integrity, and authentication of data.

By employing AES, data can be encrypted before uploading it to the cloud storage, ensuring that even if unauthorized individuals gain access to the storage, the data remains secure. The AES encryption key can be defined by the user, allowing only authorized users with the correct key to access the data. Additionally, access control mechanisms can be implemented to restrict users' access to specific data, further enhancing data security.
The utilization of AES in the Secure Deduplication with User-Defined Access Control in Cloud Storage project significantly enhances the security of the stored data, safeguarding sensitive information from unauthorized access.

B. MD5 (Message Digest Algorithm 5)

MD5 is a widely-used cryptographic hash function that plays a crucial role in secure deduplication systems in the cloud. Cloud deduplication involves identifying and eliminating duplicate data within a storage system to optimize storage space.
MD5 is a hashing algorithm that takes a text input and converts it into a fixed-length 128-bit hash code. The generated hash code will be the same for identical text inputs, but even slight changes in the text can produce entirely different hash values. This property of MD5 aids in the deduplication of files.
When employing MD5 in secure deduplication, each file is assigned a unique MD5 hash value, which is used to identify and eliminate duplicate copies of the same file. This process effectively reduces storage space in the cloud storage system.
Overall, the integration of MD5, in combination with user-defined access control, contributes to improved security and efficiency in cloud storage systems by minimizing the amount of duplicated data stored and allowing users greater control over data access.

VI. RESULT

In this section we examine the findings the outcomes of our implementation and testing. We evaluate how well the system performs in terms of deduplication efficiency, security measures, system performance, and user satisfaction. By analyzing the results, we gain a better understanding of how effectively our system provides secure deduplication with user- defined access control in the cloud. This analysis helps us identify strengths, weaknesses, andareas for improvement in our project.

A. Result Analysis

1) Secure deduplication is a process that identifies and eliminates duplicate copies of data in cloud storage, resulting in efficient utilization of storage space. By removing redundant data, organizations can significantly reduce storage costs and improve overall system performance.
2) In addition to storage optimization, secure deduplication with user-defined access control enhances data privacy and security. User-defined access control allows individuals to specify who can access their data, ensuring that sensitive informationremains protected from unauthorized access. This level of control gives users peace of mind and helps organizations comply with privacy regulations.
3) By implementing secure deduplication with user-defined access control, organizations can achieve a balance between efficient storage management and datasecurity. This approach enables them to store data more efficiently, reducing the storage footprint while maintaining the integrity and confidentiality of their information assets.
4) Moreover, secure deduplication and user-defined access control contribute to data governance and compliance efforts. Organizations can establish policies and enforceaccess controls that align with their regulatory requirements and internal data governance frameworks.

Overall, the combination of secure deduplication and user-defined access control provides a comprehensive solution for cloud storage. It not only optimizes storage space but also strengthens data privacy and security, enhances data governance, andenables organizations to meet regulatory obligations.

B. Results

1) Login Result
2) Login scenarios
a) When user is not registered in the system.
b) When user is registered, but waiting for admin activation.
c) When user is registered and successfully logged in.

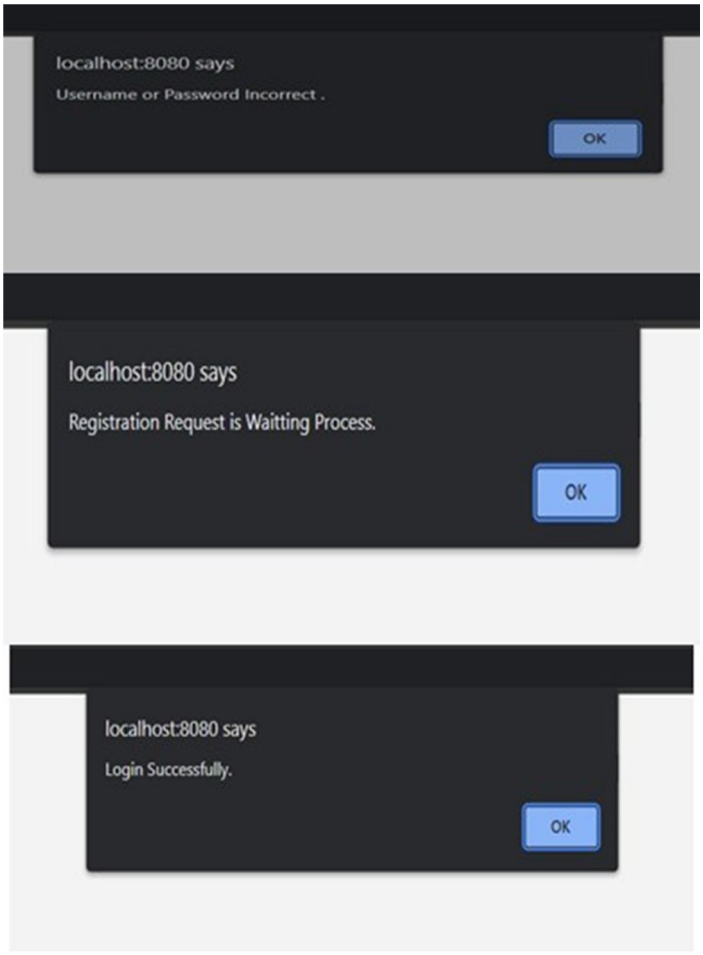


Fig. Login Popups

C. Upload File Result

When user wants to upload file, he/she need to select the file to upload. After selectingfollowing scenarios occur.

1) When file does not exist in our system.

2) When file exist in out system.

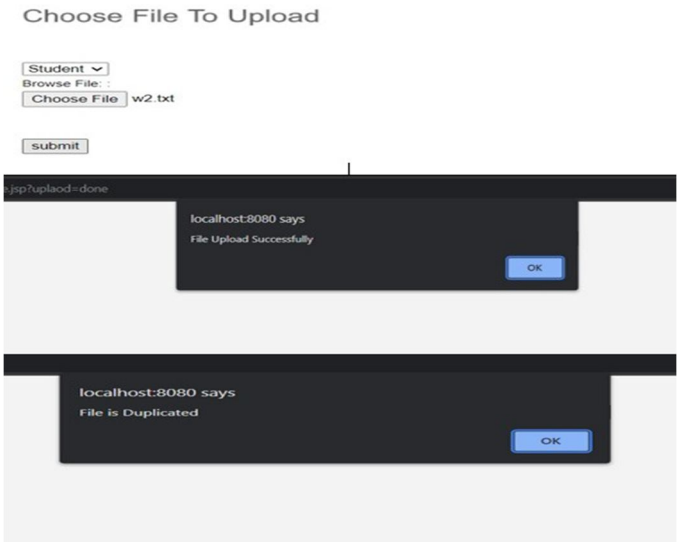
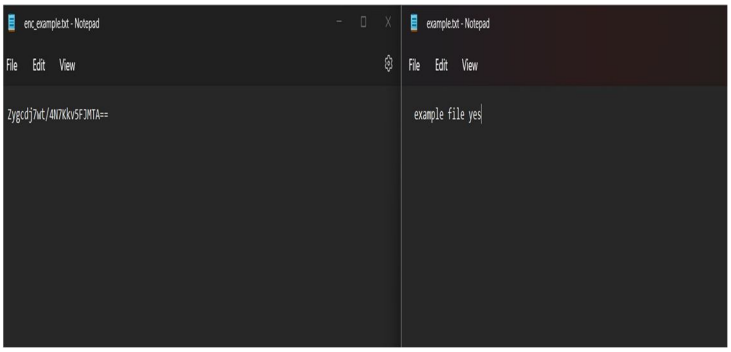


Fig. Upload File results

D. Encryption Result

Encrypted format of uploaded file is shown in the left hand side of the Fig.5.3 whileoriginal is shown on the right hand side.



Below Fig.5.4 shows that all files stored in database are in encrypted format.

u_id	File_Name	User	file
1	w2.txt	vivek	3R1JR6BFoCAaCRu3vqhc9bQj8TFGtBz8cugNpdVXzxxOHV6R2t...
2	example.txt	vivek	Zygcdj7wt/4N7Kkv5FJMTA==
3	Tushar project.doc	vivek	8NT57h2XY87OY/rx5m1uY3gJVuBSNhkS/37lox8idlAjWYcEe2...
7	swapnil.txt	swapnil	n1InZoziW6yD3nDKOB6D1V+HUM21sD3L6sy9Fdylgak=
8	ad 3.docx	vivek	jYQdSmYAFhuWSONdc/9ewqQygGOsXzyclml+SSTMAlsA7N9b5+...

Fig. Files with their encrypted format

id	key1	privilege
1	0x909ea77f312fea58c85f9c0251bebfd9	Student
2	0x38ac8fea70e0cae28c65433a63849761	Student
3	0x04c83c1c7d660925ef1386f777c772a4	Student
7	0xc2d95104673c0603a1c2246c0b60688b	Student
8	0x6f813eda0fc7442941c74ab067c2da02	Student

Fig. Generated Hash key stored in system

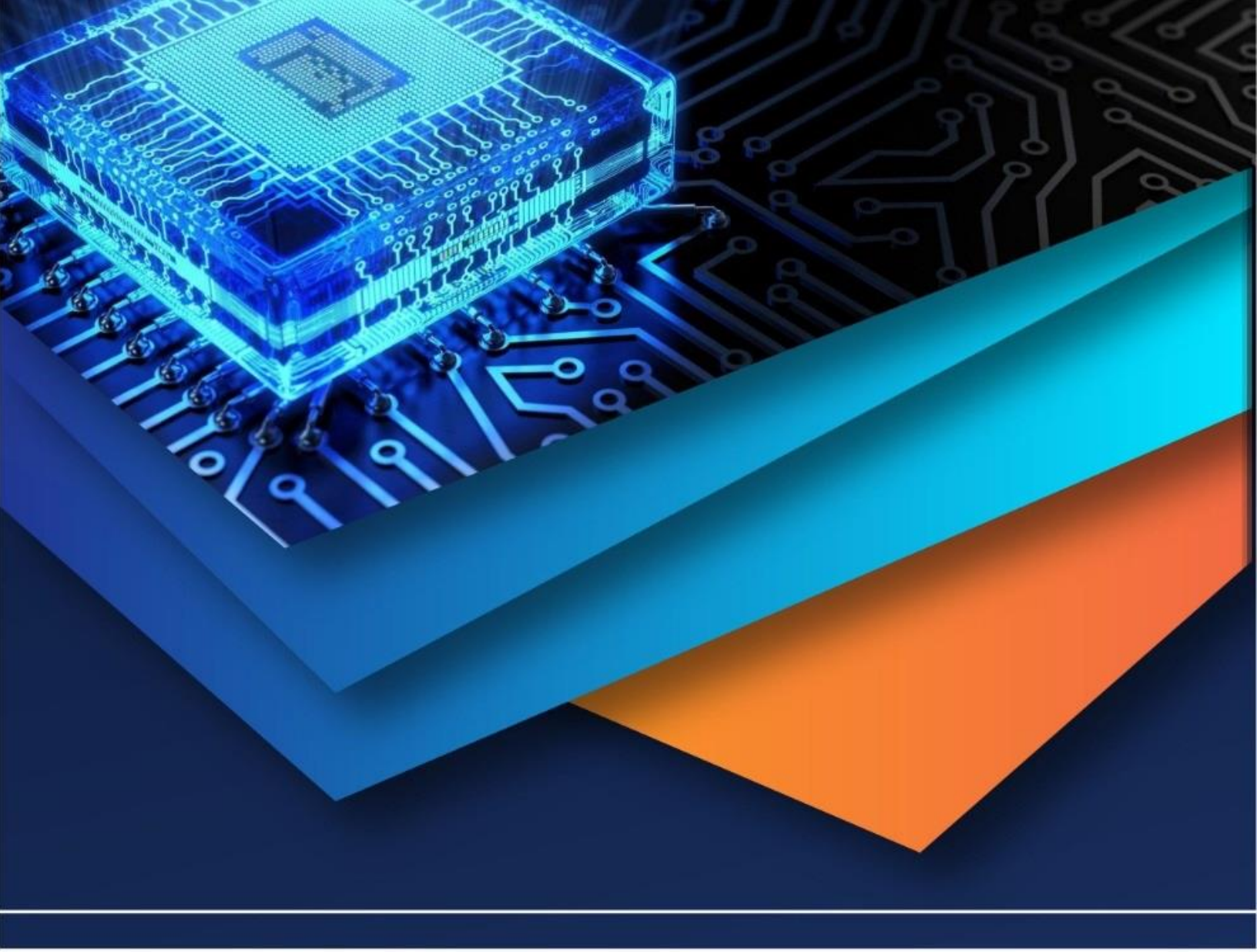
VII. CONCLUSION

The Secure Deduplication with User-Defined Access Control in Cloud Storage system aims to provide a secure and efficient deduplication technique that allows users to have control over their data access in the cloud. The system proposes an access control mechanism that leverages the attribute encryption standard (AES) and MD5 (Message Digest) algorithms to ensure data deduplication. The proposed solution has been thoroughly evaluated through simulations and experiments, and the results demonstrate its effectiveness in reducing storage overhead and enhancing data privacy while maintaining authorized access to users' data. These outcomes have significant implications for cloud storage security, offering a promising solution for secure and efficient data deduplication in cloud storage environments.

Overall, the Secure Deduplication with User-Defined Access Control in Cloud Storage system makes a valuable contribution to the field of cloud storage security. It provides a practical solution for users who seek to securely store their data and exercise control over its access.

REFERENCES

[1] https://ieeexplore.ieee.org/document/9069266 Xue Yang, Rongxing Lu, Senior Member, IEEE, Jun Shao, Xiaohu Tang, Member, IEEE, and Ali A. Ghorbani, Senior Member, IEEE 2020 Date of Publication: 16 April 2022
[2] https://www.sciencedirect.com/science/article/abs/pii/S15 74119215001248 Yanjiang Yang a,∗ , Haiyan Zhub , Haibing Luc , Jian Weng d , Youcheng Zhang e, Kim-Kwang Raymond Choo Date of Publication: 13 May 2016
[3] https://ieeexplore.ieee.org/document/7478544 Zheng Yan, Mingjun Wang, and Yuxiang Li, Xidian University, China Athanasios V. Vasilakos, Lulea University of Technology, Sweden Date of Publication: 25 May 2016
[4] https://ieeexplore.ieee.org/abstract/document/6658013 Darpan Anand; Vineeta Khemchandani; Rajendra K. Sharma Date Added to IEEE Xplore: 11 November 2013
[5] https://ieeexplore.ieee.org/abstract/document/9072400 A Novel Improvement With an Effective Expansion to Enhance the MD5 Hash Function for Verification of a Secure E-Document Date of Publication: 20 April 2020
[6] https://ieeexplore.ieee.org/document/9069266/ Achieving Efficient Secure Deduplication With User-Defined Access Date of Publication: 16 April 2020
[7] https://ieeexplore.ieee.org/document/7070725 A Practical and Effective Sampling Selection Strategy for Large Scale Deduplication Date of Publication: 27 March 2015
[8] https://ieeexplore.ieee.org/document/8936222 Secure Textual Data Deduplication Scheme Based on Data Encoding and Compression Date Added to IEEE Xplore: 19 December 2019 Secure Deduplication with User-Defined
[9] https://www.geeksforgeeks.org/advanced-encryptionstandard-aes/
[10] https://www.geeksforgeeks.org/what-is-the-md5- algorithm/
[11] https://youtu.be/G_qtQgRmiWk
[12] https://docs.oracle.com/javase/8/docs/api/java/security/MessageDigest.html
[13] https://youtube.com/playlist?list=PLrzWQu7Ajpi0RER 5EWEpScyd0NVFRQH8Y
[14] https://www.simplilearn.com/tutorials/cyber-security-tutorial/md5-



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY