International Journal for Modern Trends in Science and Technology, 8(12): 144-146, 2022

ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST0812022
Available online at: http://www.ijmtst.com/vol8issue12.html

Redundant Data Elimination Based on Secured Authorized Data Deduplication

Kancherla Rajasree

Assistant Professor, Department of Computer Science, P.B.Siddhartha Arts & Science college, Vijayawada, Andhra Pradesh, India.

To Cite this Article
Kancherla Rajasree, Redundant Data Elimination Based on Secured Authorized Data Deduplication, International Journal for Modern Trends in Science and Technology 2022, 8(12), pp. 144-146. https://doi.org/10.46501/IJMTST0812022

Article Info
Received: 26 November 2022; Accepted: 25 December 2022; Published: 31 December 2022.

ABSTRACT

An individual user has access to infinite storage space, data availability, and accessibility from any location at any time thanks to cloud computing. Although data deduplication removes redundant data and replicated data occurs in cloud environment, cloud service provider is able to maximize data storage space by incorporating data deduplication into cloud storage. Duplication of data is one of the major problems in the current scenario. Due to the increasing of data day-by-day the occurrence of more duplicate data means the availability in the cloud decreases. To avoid this, we propose a system which checks the files before they are uploaded into the cloud and uploads them if no files of same content are available in the cloud. Data privacy preservation is a crucial issue to think about in a cloud environment, and a new sort of convergent encryption technology is utilized to encrypt the data before outsourcing in order to facilitate deduplication and ensure this data secrecy. This paper describes how cloud serviceand storage providers use data deduplication without granting users access to their encrypted or plain-text data.

Keywords: Deduplication; privacy preserving, convergent encryption technique.

1. INTRODUCTION

Cloud computing is a model to provide access to computing resources and applications available on the Internet Cloud computing platform offers the network resources and storage space to the remote users. Information can be accessed by the user at any time and from anywhere via Internet. So the user and his data need not to be on same physical location. Moreover, the user does not requires to manage the actual resources. Cloud computing enables the users to access shared resources by providing services as per user requirement over the network to perform operations. Managing and deploying certain applications developed by the users can be done through cloud computing services. The use of cloud computing has widespread in the IT industry. Companies like Microsoft, Google and IBM deliver their services to its users using cloud. Because of cloud computing's high scalability and availability, it increases response time, which results in high performance and provides services to its users on a large scale.

Cloud Computing era has lots of research issues. Deduplication is one of them. It is a compression technique which identifies and locates duplicate data. It then eliminates duplicate copies of repeating data and saves the space for data that needs to be physically

store. Hence, the two main advantages of data deduplication are Reduction in Storage Allocation and Efficient Volume of Replication.

2. 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.

3. METHODOLOGY

In the proposed methodology, the main aspect is to make the data storage in cloud to be productive and secure. Here we are setting a methodology to eliminate the redundant data i.e the duplicate data. In this system initially while uploading the data to the cloud it will check for the file if it already exists and then it decides whether to store or reject similar is the case with different file types and images.

In many cases and daily activities, we are observing these things where duplicate data is occupying the storage due to that we have wastage of space to overcome this problem we have proposed this solution. Mostly in the case of cloud storage we will pay for the storage that is being used by us. So effective utilization of that space is the main aspect behind the proposed system.

4. RESULTS

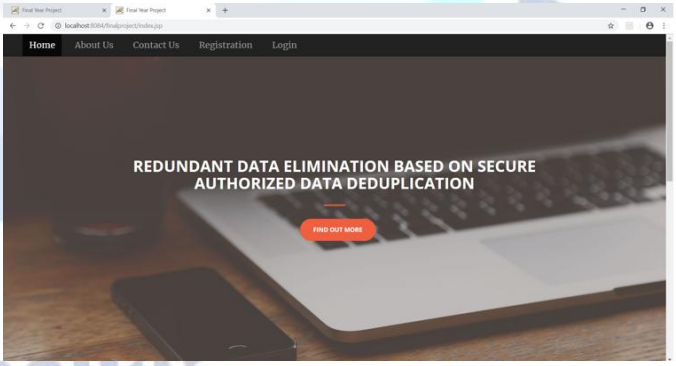

Fig : 1 showing the main page of the application

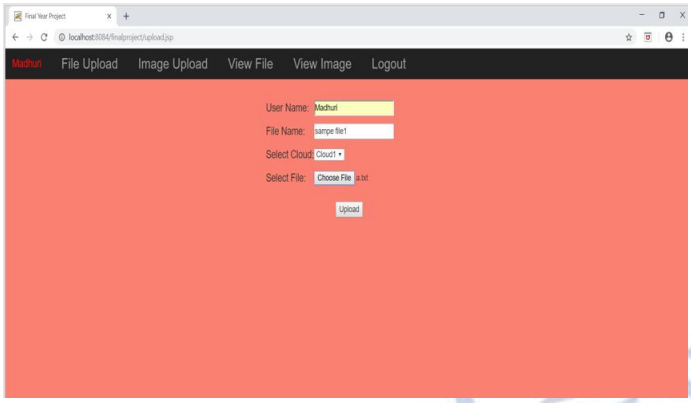

Fig :2 showing the file upload process of the application

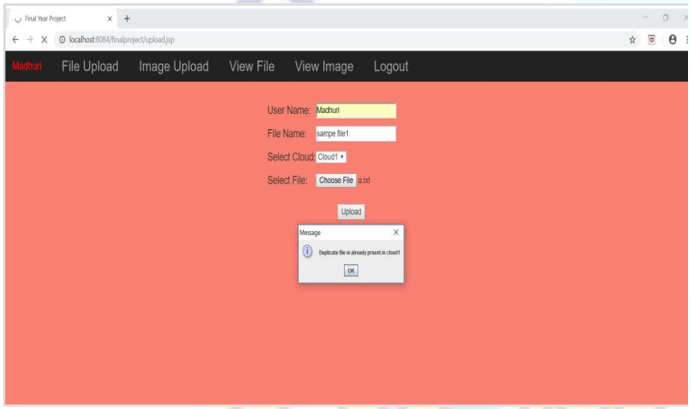

Fig :3 showing the alert that duplicate file already exists

5. CONCLUSIONS

In this system, we have proposed an effective solution to eliminate the redundant data, where files will be uploaded to cloud only if the similar file doesn't exist already in the cloud storage. The process appears like while we try to upload a file initially it checks for the similar file in the cloud storage then if it finds it then alert message will be returned like file already exists.

We have observed that what happens when we have lot of storage like lot many files are already stored it has to compare all the existing files such that it returns whether it exists or not in this context the future scope will be working on the area to reduce the time it takes.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

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/S1574119215001248 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=PLrzWQu7Ajpi0RER5EWEpScyd0NVFRQH8Y

[14] https://www.simplilearn.com/tutorials/cyber-security-tutorial/md5-