https://doi.org/10.46335/IJIES.2022.7.4.11 **e-ISSN: 2456-3463**
Vol. 7, No. 4, 2022, PP. 52-57

International Journal of Innovations in Engineering and Science, www.ijies.net

Implementation on Secure Storage for Cloud with Duplication Checking

Prof. Rahul Bhandekar [1], Prof. Pranjali Manmode [2], Juhi Bomble[3]

[1,2]Asst. Professor, [3]Student
Department of Computer science Engineering, Wainganga College Of Engineering And Management, Nagpur

bomblejuhi@gmail.com

***Received on**: 05 April, 2022* ***Revised on**: 04 May ,2022,* ***Published on**: 06 May, 2022*

***Abstract** – Cloud computing provides computer resources to consumers as a utility on demand over the Internet, and it now plays a large role in the commercial domain. Cloud storage has become one of the most popular services provided by cloud computing. Customers profit the most from cloud storage since they may save money on storage equipment purchases and maintenance by just paying for the quantity of storage they use, which can be scaled up and down based on demand. With the growing quantity of data in cloud computing, a reduction in data volumes could help providers save money by cutting the expenses of running large storage systems and conserving energy. As a result, in order to improve storage efficiency, data DE duplication techniques have been employed in cloud storage. Data usage in the cloud fluctuates over time due to the dynamic nature of data in cloud storage. Some data pieces, for example, may be retrieved often one time but not the next. Some datasets are often seen or modified by multiple people at the same time, while others require a high level of redundancy to maintain stability. As a result, having this dynamic functionality in cloud storage is crucial. On the other hand, current solutions are primarily focused on a static scheme, limiting their full applicability in the dynamic nature of data in cloud storage. For cloud storage, we offer a dynamic DE duplication approach. in this research, with the goal of increasing storage economy while maintaining redundancy for fault tolerance.*

Keywords- Data De-duplication, cloud, Advanced Encryption standard algorithm (AES), Message Digest 5 algorithm (MD5)

I- INTRODUCTION

The basic ABE system does not provide secure DE duplication, a technique for reducing storage space and network bandwidth by deleting duplicates of encrypted data stored in the cloud. Existing architectures for secure duplication, on the other hand, do not use attribute-based encryption, as far as we know. Given the ubiquitous use of ABE and safe duplication in cloud computing, a cloud storage solution that combines the two traits would be perfect. Consider the following scenario while creating an attribute-based storage system for secure encrypted data duplication in the cloud: the cloud will not store a file more than once, even if it gets multiple copies encrypted with different access permissions. In an attribute-based storage system that uses CP-ABE for data encryption, however, equipping the private cloud with such a tag checking capability is insufficient to conduct duplication. In the proposed attributed-based system, the same file could be encrypted to different cypher texts with different access policies; however, storing only one cypher text of the file means that users whose attributes satisfy the access policy of a discarded cypher text (but not the access policy of the stored cypher text) will be denied access to the data they are entitled to. To address this issue, we've introduced cypher text regeneration as a new feature in the private cloud. In terms of the adversarial model of our storage

https://doi.org/10.46335/IJIES.2022.7.4.11 e-ISSN: 2456-3463
Vol. 7, No. 4, 2022, PP. 52-57

system, we assume that the private cloud is curious but honest, in that it will try to obtain the encrypted messages but will follow the protocols faithfully, whereas the public cloud is re-elect, in that it may tamper with the label and cypher text pairs outsourced from the private cloud (note that such type of misbehaviour will be detected by the private cloud or the user via the accompanied label). Another distinction between public and private clouds is that the former cannot cooperate with users, whilst the latter may. This assumption is supported by reality, in which the private cloud is seen as more reliable than the public cloud. We assume that data users will try to gain information beyond the scope of their authorised access. In addition to attempting to steal plaintext data from the cloud, malicious outsiders may utilise duplicate phoney attacks. The system may conclude that security and performance are critical for next-generation large-scale systems like clouds. As a result, as a safe data replication challenge in this project, we will address the issues of security and performance. Users' files are judicially split into bits and replicated at crucial cloud sites in the present approach, Division and Replication of Data in the Cloud for ultimate Performance and Security. A file is fragmented depending on a set of user criteria, with each piece containing no relevant information. Each cloud node (the term node refers to computing, storage, physical, and virtual computers in this system) contains a unique fragment to increase data security.

II- OBJECTIVES

- To Establish a Safe Cloud Computing Environment.
- To develop a prototype web based application using JSP& servlet this will run as a local host using apache tomcat server.
- To develop a System that ensures reliable and secured cloud storage.
- To promote the use of (drops) for divide and replication of data in the cloud.
- To protect DE duplication in order to save storage space for cloud storage providers.

III- RELATED WORK

The RevDedup technique was proposed by Chun-Ho Ng et al. in 2013 to locate and eliminate duplicates from virtual machine pictures. When a new VM image is received. the RevDedup detects a similarity with existing data and removes it from the existing data [1]. Mihir Bellare et al. introduced a cryptographic technique called Message-Locked Encryption in the same year (MLE). The encryption and decryption keys of MLE are generated from the message itself. It was the most secure method of deduplication [2]. Zhou Lei et al. proposed a method for storing photos that employed the fixed size block method in 2014. This method generates a fingerprint directory by producing a succinct digest called a fingerprint for each image file. It generates fingerprints for new image input and compares them to a database of fingerprints [3]. In the same year, Waraporn Leesakul et al. suggest a new approach for boosting the efficiency of cloud storage capacity by using dynamic data deduplication. This strategy preserved redundancy while increasing storage space [4]. Zhou Lei et al. proposed a method for storing photos using the fixed size block method in 2014. This method generates a fingerprint directory by producing a succinct digest called a fingerprint for each image file. It generates fingerprints for new image input and compares them to a database of fingerprints [5]. In their survey on security concerns in clouds and security solutions, Issa M. Khalil et al. discovered 28 cloud security vulnerabilities [6]. in the same year. N. Jayapandian et al. introduced the authorization-based scheme in 2015. This system uses differential rights based on duplicate check to safeguard user data confidentiality [7]. Mi Wen et al. devised a secure deduplication strategy employing convergent encryption technique in the same year [8]. Lakshmi Pritha et al. built a system that uses RSS keys to offer safe access to cloud resources and demonstrated the ALG technique for data deduplication the same year [9]. Chun-I Fan et al. presented a check block approach for encrypted data deduplication the same year [10]. Mr. Dame Tirumala Babu et al. presented a solution for data deduplication based on authorization to protect data the same year [11]. In 2016, Shuai Wang and colleagues proposed the RRMFS file system to help in data deduplication. [12]. The following year, Zheng Yan et al. proposed a mechanism for ownership and reencryption to deduplicate encrypted data stored in the cloud [13]. The destor tool was used by Naresh Kumar et al. in the same year to compare numerous deduplication algorithms. In the data deduplication strategy [14], fixed length and variable length chunking techniques are used. Jun Rene et al. published a safe data deduplication solution based on differential privacy the same year [15]. In the same year, Saurabh Singh et al. published a cloud security survey which included a discussion of security issues and challenges[[16]. The Load Balanced Flow Scheduling technique for dynamic load balancing and network performance maximisation was introduced by

https://doi.org/10.46335/IJIES.2022.7.4.11 **e-ISSN: 2456-3463**
Vol. 7, No. 4, 2022, PP. 52-57

International Journal of Innovations in Engineering and Science, www.ijies.net

Feilong Tang et al. in the same year [17]. Danoing Li et al. presented the CSPD strategy to improve duplicate check accuracy in 2017 [18], which uses a modified DCT-based Perceptual Image Hash (D-phash). In the same year, Hui Cui et al. developed an attribute-based ABE encryption solution for cloud storage. In the same year, Rayan Dasoriya et al. [20] published a dynamic load balancing technique that balanced the load across multiple connected network links. In the same year, Jiang et al proposed a data secrecy and ownership management system for data deduplication based on Proof of Ownership (PoW) [21],In the same year, Himshai Kambo et al. developed a secure deduplication system based on the CDC and MD5 algorithms. The MD5 algorithm produced hash values for the segments or chunks created by CDC, which were fragmented using randomness. It was used in order to boost network bandwidth [22].

IV- PROPOSED SYSTEM

An attribute-based storage system facilitates secure de-duplication. Our storage solution is built on a hybrid cloud architecture, which means that a private cloud manages computing and a public cloud manages storage. Thanks to an attribute-based storage system that provides secure de duplication of encrypted data in the cloud, even if it receives several copies of the same file encrypted under differing access permissions, the cloud will not store a file more than once. The Attribute Authority assigns each user a decryption key that is tied to a set of characteristics. File duplication is checked using the attribute-based storage system. The file is saved without being duplicated. If there is a duplication, the attribute authority changes the ownership permission. In this method, client accreditations are utilised to validate the client's confirmation. There are two types of cloud available in such situations: private cloud and open cloud. The private cloud stores the client accreditation, while the open cloud displays the customer information. The system used a half-and-half cloud formation paradigm as part of the suggested system. The file name must be remembered when duplicating records in this method, and information DE duplication is checked at the square level. If a client wishes to retrieve information or download a record, however, he must first download both documents from the cloud server, which will result in the operation being conducted on the same record. This compromises the security of distributed storage. DROPS (Division and Replication of Data in the Cloud for Optimal Performance and Security) is a methodology for dealing with cloud security and performance issues. In this project, the DROPS technique is utilised to split a file into pieces and replicate the fragmented data among cloud nodes. Each node only keeps a single piece of a specific data file, ensuring that no meaningful information is revealed to the attacker in the event of a successful attack.

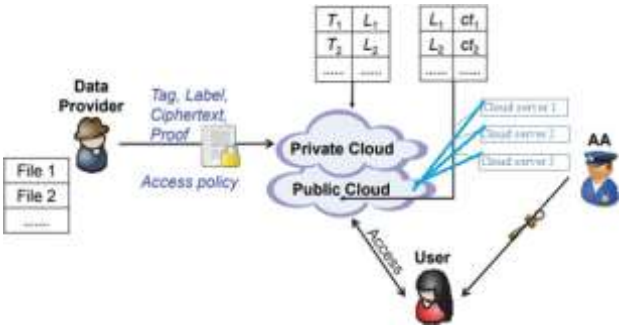

Fig 1- Proposed System

Diagram description

- The data user will upload the information, which will be verified for duplicates before being uploaded to the cloud.
- Before uploading to the cloud, the file will be broken into pieces and duplication will be verified at the chunk level.

.Software installation

• Installation of JDK 1.8

1. To run the installation application, double-click jdk-8-ea-bin-b32-windows-i586. The JDK License dialogue box appears. Accept the licence in order to install JDK.
2. You can pick a custom directory for JRE Files in the JRE Custom configuration dialogue.
3. A successful installation is indicated by the dialogue.

• Installation of Net Beans IDE 7.3.1 To set up the software, follow these steps:

1. Run the installer after the download is complete. The installation executable file for Windows has the.exe extension. Click the installer to start dubble.
2. Customize your installation if you downloaded the whole thing. At the installation wizard's welcome page, do the following: a. Click the Customize button.

https://doi.org/10.46335/IJIES.2022.7.4.11 **e-ISSN: 2456-3463**
Vol. 7, No. 4, 2022, PP. 52-57

International Journal of Innovations in Engineering and Science, www.ijies.net

Make your selections in the Customize Installation dialogue box. Click Next on the installation wizard's welcome page. Get a review of the licence agreement on the License agreement page, then click the Accept check box and Next. Decide whether you wish to install JUnit or not on the JUnit License Agreement screen, then select the appropriate option and click Next. Perform the following steps on the NetBeans IDE installation page: Accept the default NetBeans IDE installation directory or select a different path. Note that the installation directory must be empty, and the user profile you're using to run it must have read/write access on it. Accept the JDK installation that comes with the NetBeans IDE by default. If the installation wizard was unable to locate a compatible JDK installationYour JDK isn't in the correct location. In this instance, click Next or cancel the current installation and indicate the path to be installed. You can resume the installation after installing the appropriate JDK version. If you're installing Apache Tomcat, accept the default installation location or provide a different path on the installation page. Next should be selected. Verify that the list of components to be installed is correct on the Summary page, and that you have enough space on your system for the installation. To begin the installation, click Install. Provide unidentified consumption data if desired on the Setup Complete page, then click Finish.

- MySQL Database

Microsoft SQL Server is a database management system that was developed by Microsoft. It serves as a database server, storing and retrieving data for other software applications. This can be executed on the same computer or on a networked computer (including the Internet). Microsoft SQL Server is available in a variety of editions aimed at different audiences and workloads ranging from modest single-machine applications to huge Internet-facing applications with many concurrent users.

Algorithms

1) The MD5 (message-digest algorithm)

Algorithm is followed by the steps below.

Step 1: Attach the Padding Bits. The message is "padded" (stretched) to a length (in bits) of 448 modulo 512.

Step 2: Add the Length...

Step 3: Initialize the MD Buffer

Step 4: Break down the message into 16-word chunks.

Step 5: Produce.

MD5 (Message-Digest algorithm 5) is a cryptographic hash function with a 128-bit length that is widely used in cryptography. As an Internet standard (RFC 1321), MD5 has been utilised in a wide range of security applications, and it is frequently used to check the integrity of data. An MD5 hash is represented by a 32-digit hexadecimal number.

2) AES(Advanced Encryption standard)

The Advanced Encryption Standard (AES) is the most popular and commonly used symmetric encryption method presently (AES). It is discovered six times more quickly than triple DES.Because the key size of DES was too small, it needed to be replaced. It was designed to be vulnerable to a key search attack. Triple DES was designed to solve this disadvantage, but it was slow.

AES's characteristics

- 128-bit data, 128/192/256-bit keys
- Symmetric key symmetric block cypher
- Provide comprehensive specification and design information
- Software implementable in C and Java
- Stronger and quicker than Triple-DES

The illustrative of AES structure are as follows–

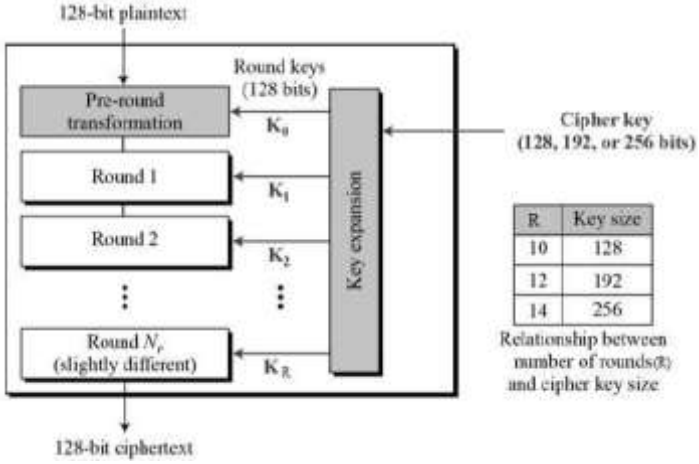

Fig 2 . AES Structure

Encryption process

We are now down to the AES encryption description round. There are four sub-processes in each round. The first-round decryption process is shown below.

https://doi.org/10.46335/IJIES.2022.7.4.11 e-ISSN: 2456-3463
Vol. 7, No. 4, 2022, PP. 52-57

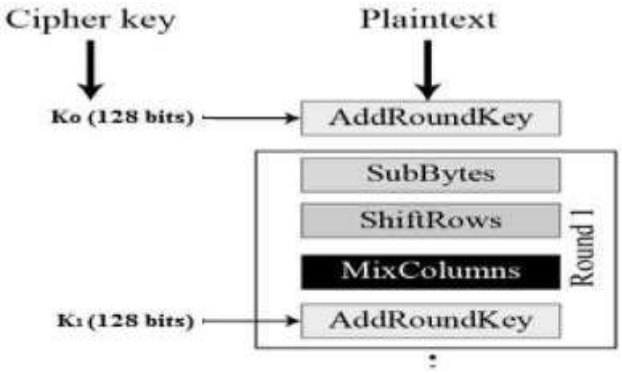

Fig 3- Encryption flow

Decryption process

The reversed-order decryption of an AES ciphertext is identical to the reversed-order encryption method. Each round includes all four processes.

• Add round key

• Mix columns

• Shift rows

• Byte substitution

As a result, the encryption and decryption algorithms are implemented individually in each round, despite the fact that they are closely related.

Pseudo code

state =M
AddRoundKey(state,&w[0])
For i=1 step 1 to 9
SubBytes(state)
ShiftRows(state)
MixColumn(state)
AddRoundKey(state, &w[i*4])
end for
SubBytes(state)
ShiftRows(state)
AddRoundKey(state,&w[40])

AES Analysis

AES is widely used in both hardware and software in modern cryptography. No viable AES attacks have been devised to date. AES also has the capacity to execute exhaustive key searches because of its key length flexibility.

V- ADVANTAGES

- **Confidentiality of data**
  We provide data confidentiality in this case, which entails encrypting data to prevent unauthorised access to data and ensuring that only authorised users have access to it.
- **Data Security:**
  Using a cryptographic solution, we can keep the content of data hidden, such as sensitive information, and keep our data safe from unwanted users, preventing data leakage.
- **Controlling access**
  Access control is a technique for granting or denying access to a system. A client-side depulication proposal allows for the forward and backward security of outsourced data to be controlled by a single user.
- **Data Backup on the Cloud**
  Cloud backup allows businesses and individuals to save data from their computers in the cloud. The data backup procedure begins when the data owner requests the data that was previously stored in the cloud.

VI- RESULT

UML Module

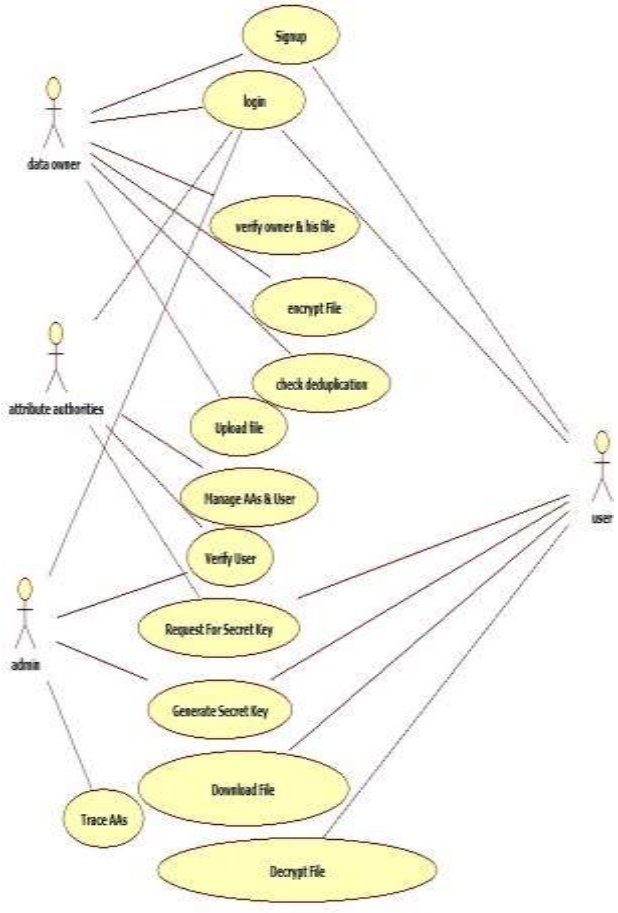

Fig 4-Use Case Diagram

Module description

User module

In this module, the user must first register and obtain a secret key, which they will use to login. They will then be able to retrieve the client's submitted files as well as view anything saved in the cloud.

VII- CONCLUSION

As a result, we'll be able to develop a secure cloud deduplication solution. The files will be checked to determine whether they've been uploaded before, and if they have, they won't be. The efficiency of the cloud storage system will be improved with this approach. It will, to a considerable degree, solve the problem of storage space scarcity.As a result, we'll create a web-based prototype model for cloud-based secure storage with duplication detection. This system will stop duplicate files from being uploaded to the cloud. It liberates a significant amount of cloud space.

REFERENCES

[1] *L. Meegahapola, N. Athaide, K. Jayarajah, S. Xiang, and A. Misra, "Inferring Accurate Bus Trajectories from Noisy Estimated Arrival Time Records," 2019 IEEE Intell. Transp. Syst. Conf. ITSC 2019, pp. 4517–4524, 2019, doi: 10.1109/ITSC.2019.8916939.*

[2] *A. Bhalerao and A. Pawar, "Utilizing Cloud Storage for Big Data Backups," pp. 933–938, 2018.*

[3] *L. Araujo, "Genetic programming for natural language processing," Genet. Program. Evolvable Mach., vol. 21, no. 1–2, pp. 11–32, 2020, doi: 10.1007/s10710-019-09361-5.*

[4] *H. Cui, R. H. Deng, Y. Li, and G. Wu, "Attribute-based storage supporting secure deduplication of encrypted data in cloud," IEEE Trans. Big Data, vol. 5, no. 3, pp. 648–660, 2019, doi: 10.1109/TBDATA.2017.2656120.*

[5] *H. Hou, J. Yu, and R. Hao, "Cloud storage auditing with deduplication supporting different security levels according to data popularity," J. Netw. Comput. Appl., vol. 134, pp. 26–39, 2019, doi: 10.1016/j.jnca.2019.02.015.*

[6] *https://aws.amazon.com/what-is-cloud-computing/*

[7] *Dr.P.Sujatha and Dr.P.SriPriya, "Security Threats and Preventive Mechanisms in Cloud Computing ", JASC: Journal of Applied Science and Computations Volume V, Issue XII, December/2018 ISSN NO: 1076-5131.*

[8] *K.Sharmila S. Borgia Anne Catherine Sreeja V.S, "A comprehensive Study of Data Masking Techniques on cloud", International Journal of Pure and Applied Mathematics Volume 119 No. 15 2018, 3719-3727.*

[9] *N. Lakshmi Pritha and N.Velmurugan, "Deduplication Base Storage and Retrieval of Data from Cloud Environment" in Conference on Innovation Information in Computing Technologies, Chennai, pp. 1-6, IEEE 2015.*

[10] *Chun-I Fan and Shi-Yuan Huang, "Encrypted Data Deduplication in Cloud Storage", Article in 'ASIAJCIS' 15 Proceedings of the 2015 10th Asia Joint Conference on Information Security, pp.18-25, May 24-26, 2015, IEEE Computer Society, Washington, ISBN: 978-1-1989-5.*

[11] *Dama Tirumala Babu and Yaddala Srinivasulu, "A Survey on Secure Authorized Deduplication Systems", International Research Journal of Engineering and Technology. Volume: 02 Issue: 05. Aug-2015.*

[12] *Shuai Wang and Jianhai Du "A Storage Solution for Multimedia Files to Support Data Deduplication", 2016 2nd International Conference on Cloud Computing and Internet of Things, Dalian, China, pp-78-8, 2016.*

[13] *Zheng Yan and Wenxiu Ding," Deduplication on Encrypted Big Data in Cloud", IEEE Transactions on Big Data, Vol. 2, No. 2, April-June, 2016.*

[14] *Naresh Kumar, Preeti Malik, Sonam Bhardwaj, Sushil Chandra Jain, "Comparative Analysis of Deduplication Techniques for Enhancing Storage Space", 4th International Conference on Parallel, Distributed and Grid Computing. IEEE, 2016.*

[15] *Jun Ren and Zhiqiang Yao, "A Secure data deduplication scheme based on differential privacy", IEEE 22nd International Conference on Parallel and Distributed System, pp-1241-1246, 2016.*

[16] *Dr VK Govindan and BS Shajee Mohan. Idbe–an intelligent dictionary based encoding algorithm for text data compression for high speed data transmission. In Proceeding of International conference on Intelligent signal processing, 2004.*

[17] *deduplication overview. https://docs.microsoft.com/enus/windows-server/storage/data-deduplication/overview. Accessed on August 2017.*

[18] *Qin Jiancheng, Lu Yiqin, and Zhong Yu. Fast algorithm of truncated burrows-wheeler transform coding for data compression of sensors. Journal of Sensors, 2018, 2018.*

[19] *Sriram Keelveedhi, Mihir Bellare, and Thomas Ristenpart. Dupless: server-aided encryption for deduplicated storage. In Presented as part of the 22nd {USENIX} Security Symposium ({USENIX} Security 13), pages 179–194, 2013.*

[20] *Juha Kivij¨arvi, Tiina Ojala, Timo Kaukoranta, Attila Kuba, L´aszl´o Ny´ul, and Olli Nevalainen. A comparison of lossless compression methods for medical images. Computerized Medical Imaging and Graphics, 22(4):323–339, 1998.*

[21] *SR Kodituwakku and US Amarasinghe. Comparison of lossless data compression algorithms for text data. Indian journal of computer science and engineering, 1(4):416–425, 2010.*

[22] *M. O. Kulekci. A method to ensure the confidentiality of the compressed data. In 2011 First International Conference on Data Compression, Communications and Processing, pages 203–209, June 2011.*