DATA SECURITY AND ACCESS CONTROL DATA DEDUPLICATION ACROSS MULTIPLE CLOUD FRAMEWORKS

[#1]**Mrs.PINGILI SHANTHI,** *Assistant Professor*
[#2]**Mrs.MARYADA MAMATHA,** *Assistant Professor*
Department of CSE,
SREE CHAITANYA INSTITUTE OF TECHNOLOGICAL SCIENCES, KARIMNAGAR, TS.

ABSTRACT—
This paper investigated the application of a powerful method for eliminating similar data in a multi-cloud setting. It is obvious that using a private cloud as a middleman to find duplicate data is a secure method. Our approach to data sharing and storage ensures that access to the data on the public cloud is restricted to authorized users only. Data stored on the public cloud is neither encrypted or secure because of the outdated technology used.Our creative and reliable convergent key technology allows redundant data to be safely removed in a multicloud environment, which effectively tackles these challenges. Cloud servers that are open to the public house private cryptographic keys and sensitive personal information.Using this method, you can successfully remove files that are no longer needed, freeing up storage and internet space.

KEYWORDS*—Public Cloud, Convergent Key Encryption, File level Check, Block Level Check, Convergent key.*
DOI Number: 10.48047/nq.2022.20.22.NQ10502 **NeuroQuantology2022;20(22): 4937-4942**

I. INTRODUTION

Cloud computing is a means of offering information technology services. Users can now remotely access tools via a web-based application and interface, removing the need for a physical computer. Storing vast amounts of data on the cloud simplifies management and retention while reducing bandwidth and storage capacity needs. Deduplication is a well-established technique for ensuring the dependability and scalability of cloud data management. The term "de-duplication" refers to the process of removing redundant data from cloud storage in order to maximize bandwidth and storage capacity. This method improves the scalability of network data flows by lowering the amount of data to be transported and assuring continuous storage utilization. Data minimization is possible at both the block and file levels. Deduplication can help you get rid of superfluous data that has been replicated many times. Keeps the collection or file in its original state. This technology enables the reuse of empty storage space and decreases data transmission over a network. Data compression raises the risk of unauthorized access to sensitive customer information by anyone within or outside of a company who do not have the requisite authorization. As a result, I am concerned about my personal safety and privacy. Several cloud storage services, including Dropbox, Mozy, and Memopal, have begun removing obsolete user files to make room for new ones. Data movement aggravates security and privacy flaws [3]. Compression and integrated key management work together to improve performance and functionality. Standard encryption requires each user to have a unique key to protect their information in order for it to be effective. It is critical to ensure that all persons with access to the disputed file can indisputably verify ownership. Individuals with the necessary knowledge can establish a constant connection with a server, eliminating the need to relinquish said information upon presentation of evidence..

II .RELATED WORK

This essay provides a comprehensive analysis of the prior studies conducted on cloud privacy and security. We also examine other studies that exhibit certain resemblances to ours but pursue other goals.

DupLESS Server Aided Encryption for Deduplicated Storage

DupLess: Dropbox, Mozy, and other cloud storage services employ server-assisted encryption to reduce the need for duplicate storage. This strategy aims to optimize storage capacity by preserving only one copy of each file.Customers who cannot access their secured files can employ message lock encryption as an extra precautionary measure. The main goal of this system is to provide a secure and deduplicated storage solution that is resistant to brute-force attacks. Clients acquire message-based keys from a key server by utilizing an oblivious PRF protocol on a server that eliminates duplicate entries. These keys are then employed in the process. Customers have the option to safely store their data utilizing a pre-existing service, with the seller taking care of the deduplication process. Stringent privacy measures are still in effect. Utilizing encryption for deduplicated storage provides equivalent performance and space advantages as using a storage service with unencrypted data, making it a logical choice [2].

Characteristic:

1. Additional safety safeguards were put in place.
2. a simple encryption method used to eliminate redundant data.
3. Simple: A command-line tool that is compatible with both Dropbox and Google Drive would be the best option.
4. Methods for assuring the security of messages. To maintain confidentiality and security, data is encrypted using cryptography.
5. The ownership of remote-file devices must be established.

It just preserves one duplicate of the copied data. Client-side deduplication aims to detect deduplication opportunities at the client level in order to conserve bandwidth by avoiding the need to transfer duplicate files to the server[11].To respond to the attacks by Shai Halevi, Danny Harnik, Benny Pinkas, and Alexandra Shulman-Peleg, the authors propose a Proof of Ownership approach. Using this method, a client can demonstrate to a server that they genuinely own a file rather than simply providing some general information. The proposed improvements, which rely on Merkle trees and specific encodings, are thoroughly tested for security.

Characteristic:

1. to learn about the kind of security issues that client-side deduplication can guard against.
2. Ownership verification is one of the strict security measures.
3. Petabyte-scale data storage systems must adhere to tight criteria in order to function successfully.

A Secure Deduplication with Efficient and Reliable Convergent Key Management

Data deduplication is a typical approach used by cloud storage businesses to reduce the number of duplicate copies of data. This reduces the demand for storage space and upload bandwidth. Although the concept appears to be sound, there are numerous challenges that must be overcome before secure deduplication can be used in cloud storage. Convergent encryption has been extensively used in secure deduplication. Convergent encryption is still struggling to handle a large number of convergent keys consistently and efficiently.

Various approaches1.Key management

2.SConvergent Encryption [4]

Twin Clouds: An Architecture for Secure Cloud Computing

The. Sadeghi, T. Schneider, S. Bugiel, and S. Nurnberger developed a secure protocol for transmitting data and tasks across an unreliable commodity cloud. Users must pay to use cloud services in order to have consistent and ongoing access to these services. This technology verifies and preserves cloud-stored data while limiting the dangers associated with unpredictable online activity.The one cloud is in charge of efficient and secure setup activities, while the other controls sensitive data and responds to requests about outsourced data. This is how labor is distributed.

Security and Privacy issues in the cloud:

Only authorized persons should have access to data stored in the cloud. What is the significance of precise user registration for a security system that records and analyzes cloud data usage? The user must confirm that the cloud provider has implemented suitable security measures to protect their data and limit access to only those who are allowed.The most successful and extensively used technique is deduplication. Nonetheless, widespread use of cross-user compression can risk privacy dramatically. Basic technologies are capable of facilitating cross-user compression, lowering the likelihood of data loss.

III. BACK GROUND

Previously, systems that eliminated duplicates could not independently validate permissions. Nonetheless, these tests are critical in today's world. At the time of installation, this authorized deduplication system grants each user a distinct set of privileges. Check out this quick overview of diminishing cloud cover:

Deduplication

Block-level and file-level data deduplication are two popular ways. The amount of data that must be processed impacts the effectiveness of each method. Block-level reduction [27] allows you to change or keep the size of a data block. "Target-based deduplication" refers to data replication that occurs at the destination, whereas "source-based deduplication" occurs at the client's end. When forming a group, one considers the location of the reduction process. To locate and delete duplicates, the client utilizes a hashing mechanism to thoroughly check each data fragment before uploading. In order to establish communication, the client sends the generated hashes to the storage source. This method is also known as "source-based deduplication." If this method is used, the customer will only be able to input real data in discrete data segments. While client-side deduplication reduces bandwidth utilization, it exposes the system to side-channel attacks [19], which allow hackers to determine the rate at which specific data is stored. The storage service defended the system from side-channel assaults by eliminating unnecessary data. However, this strategy has no effect on transmission overhead..

Convergent Key Encryption

Convergent encryption is used to safely store deduplicated data to improve security [5]. Before encrypting the clone, the user must obtain a convergent key from each unique original instance of the object. The user can add a tag to the copied data to assist distinguish related items. Because the two sets of data are comparable, they will have the same tag if the tag is accurate. Before looking for duplicates, the human must send the tag to the computer.The individual has made a contribution of

IV. SYSTEM STUDY

Presented System:

Our approach eliminates unnecessary data at the service provider level without the need for user authentication. The information and corresponding access keys are stored at the cloud server level. The lack of sufficient security measures in cloud server access poses a significant risk, akin to employing an inappropriate and all-encompassing approach at the client level. The following ways may potentially furnish information to potential assailants. Duplication refers to the act of replicating data at the block level, while maintaining the original file name or content. Prior to initiating the data compression process, the cloud provider must ensure the establishment of Proof of Ownership (POW) and the consistent maintenance of label tags.

Disadvantages:

Inadequate measures for ensuring personal safety

Insufficient measures are taken to adequately secure personal information. without implementing measures to safeguard the data

They duplicated the data using a method that made it vulnerable to unauthorized access. Eliminating superfluous data The systems are unable to utilize differentiated authorization due to their inability to perform duplication verification.

The Proposed System

The primary objective of the proposed technique is to effectively and reliably eliminate superfluous data in a secure manner. We have developed a dependable data compression method that can discern between sensitive and non-sensitive data in real-time during its transmission to the cloud. Sensitive data is safeguarded by the use of cryptographic methods. We can confidently guarantee the information's security and confidentiality by adopting this strategy.

System Architecture

Fig 1. Proposed System Architecture

V SYSTEM IMPLEMENTATIONS.

User: Prior to uploading data to the cloud, users are required to complete the registration process by providing their name, email address, password, and mobile number.

Data Owner: - In order to register, the individual who has the data must complete the registration form on our website. By just giving their login credentials, users can access our cloud server and do tasks such as uploading or retrieving data. Data security will be guaranteed by the utilization of access controls and file encryption. Consequently, new users will be required to submit a distinct and unambiguous request in order to obtain access to the data. Verified

users will be sent an email containing their login credentials.

Data replication is implemented as an extra security measure. In the event that the public cloud identifies duplicate data, a warning notification will appear, prompting the user to upload the file again. In order to make alterations prior to publication, the user must revert back to the previous iteration. Users are provided with the choice to classify particular content as either sensitive or non-sensitive before to uploading it to a publicly accessible cloud. In addition, they can employ techniques to safeguard exclusively the data that they deem to be confidential.Any illicit entry into the object, such as when endeavoring to retrieve files, must be promptly notified to the owner.

Encryption of Files

To encrypt and decode the information in this line, a shared secret key, designated by the letter k, will be utilized. It may change back and forth between encrypted and plain text. In this case, three critical responsibilities were met:

KeyGenSE: The key generation procedure (k) returns o when the security parameter is set to 1..

EncSE (k, M):C employs symmetric cryptography. Following encryption, the message M and the secret key are used to generate the ciphertext C.

DecSE (k, C): If you enter the secret key along with the encrypted text C, function M will display the original message M.

(a)**Confidential Encryption**

Implementing procedures to avoid the retention of duplicate data safeguards data privacy. The first set of information is used to generate a unique key for each individual. Following that, the information copied is encrypted with these encryption keys. To make the process of finding duplicates easier, the user adds a unique identification to each duplicate copy of the data.

Fig 2. Confidential Encryption

(b) Proof of Data

Users can establish ownership of data stored in the system by submitting physical copies of the data [11]. Both the user and the storage service must collaborate in order to establish custody.

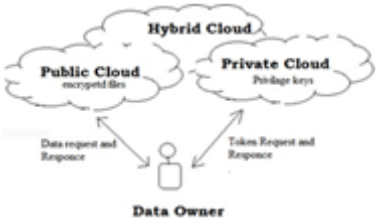

Fig 3. Proof of Data Owner

(c)Identification Protocol

1. The technique has two sequential steps that must be done in the following order:
2. An individual might engage in a certain action to verify their identity to a verifier.
3. It is critical that the verifier uses publicly available data to ensure the accuracy of the information..

Data De-Duplication Types

A.File-level de-duplication

This type of storage is alternatively referred to as file-level data storage or single-instance storage. The method verifies whether the file needs to be stored and if there is an existing copy by comparing all of the file's attributes to an index. The term used to describe this approach is de-duplication. Only the indices of distinct items will be maintained and updated. If the file is not regularly updated, a hyperlink to it will be retained at its original location. Ultimately, just a single instance of the file is preserved, and any subsequent duplicates are substituted with a "stub" that establishes a connection to the original..

B.Block-level de-duplication

Comprehensive information for every individual block By following this procedure, you will eliminate any other subfiles. By partitioning the file into smaller segments according to this criterion, we may examine the content to see if it has been previously stored or documented. Assigning a distinct identity to each individual data element is advantageous as it facilitates the process of locating associated data. The hash method assigns a distinct identity to each item of data. We will examine the disparities between the central index and the distinctive appellation that was bestowed. However, if the ID is already in use, the information has only been saved and processed once.Consequently, the sole means of accessing the primary account is by utilizing a hyperlink. If the ID is both novel and does not already exist, the block is considered to be unique. The Index's distinctive attribute has been revived, however it remains unaltered.

Various suppliers and dealers have varying degrees of verification requirements that must be met before they authorize a payment. Various systems may utilize

blocks of varying sizes. Some individuals may utilize static widths, however others may employ dynamic widths. The situation is further complicated by the possibility of individuals altering the conventional dimensions of the blocks. You have the option to select from various predetermined file sizes, and the block sizes vary between 8 KB and 64 KB. The primary distinction lies in the fact that smaller content fragments are more prone to accurately identifying themselves as duplicates. There is compelling evidence that reducing data storage leads to decreased data retention. If a file is modified and the deduplication algorithm detects a data segment that matches another segment, it may not be able to identify the duplicated data segment due to the use of fixed-size blocks, which complicates distinguishing between them. This is because any movement or alteration of the file blocks could potentially impact subsequent readings.

Restrict From Unauthorized Access

Prior to accessing another individual's data on the public cloud, users must obtain explicit consent from the data owner. Also, they can only receive privileged access keys through safe channels like email, unless the owner of the material gives them special authorization. The authorization process effectively stops illegal access by making sure that only authorized users are granted admission.

The method that is being employed right now

Before being uploaded to a public cloud, data must be encrypted by the owner of the data using a robust secure mechanism such as HMAC-OTP. This makes it safe to look at and get to users' data that is kept in public clouds. With this sort of symmetric encryption, you only need one key to encrypt and decrypt data. The outcome is better speed and easier key management. The cloud provider and hackers will not be able to view the encrypted data in this instance.

HMAC-BASED ONE-TIME PASSWORD ALGORITHM

K be a secret key, C be a counter

The HMAC is displayed as follows: HMAC(K,C) = SHA1(K 0x5c5c... SHA1(K 0x3636... C)). It was discovered using the cryptographic hash tool SHA-1. The HMAC result must consist of four bytes that can be retrieved appropriately using the truncate approach. In arithmetic notation, the exact formula is HOTP(K,C) = 0x7FFFFFFF + Truncate(HMAC(K,C)). Use a bitwise operation on the mask 0x7FFFFFFF to remove the value's highest-order bit. This prevents several difficulties from occurring if the processor—specifically, a specific processor—takes the result as a signed number.This person has not made any text posts.Before it can be used in a system, the output must be converted into a HOTP value. A HOTP number is either a six- or eight-digit code, depending on the situation.

To obtain the HOTP-Value, divide the remainder of HOTP(K,C) by 10d. In this scenario, K is the secret key and C is the counter number. HOTP should be used in conjunction with an authentication service to improve how users' identities are checked in a system. Users can also obtain a new password (OTP) value from the server and compare it to their code to ensure they are who they claim to be with the validation server.

VI. CONCLUSION

By removing duplicates, data compression ensures that each piece of information is unique. To ensure that only authorized users may access the system, we deploy convergent encryption and allow double verification. De-duplication is a technique for keeping networks from becoming overcrowded and making the most of storage space. As a result, no irrelevant or worthless data will be saved on the cloud. Someone who is not supposed to be there cannot obtain information from a source where they are not supposed to be. It has numerous advantages, including confirmed copy verification, cloud storage, privacy, and the possibility to delete recovery data.

REFERENCES

[1] P. Anderson and L. Zhang. “Fast and secure laptop backups with encrypted de-duplication”. In Proc. of USENIX LISA, 2010.

[2] M. Bellare, S. Keelveedhi, and T. Ristenpart. “Dupless: Server aided encryption for deduplicated storage”. In USENIX Security Symposium, 2013.

[3] PasqualoPuzio, Refik Molva ,MelekOnen ,”CloudDedup: Secure Deduplication with Encrypted Data for Cloud Storage”, SecludIT and EURECOM,France.

[4] Iuon –Chang Lin, Po-ching Chien ,”Data Deduplication Scheme for Cloud Storage” International Journal of Computer and Control(IJ3C),Vol1,No.2(2012)

[5] Shai Halevi, Danny Harnik, Benny Pinkas,”Proof of Ownership in Remote Storage System”, IBM T.J.Watson Research Center, IBM Haifa Research Lab, Bar IIan University,2011.

[6] M. Shyamala Devi, V.VimalKhanna,Naveen Balaji ”Enhanced Dynamic Whole File De-Duplication(DWFD) for Space Optimization in

Private Cloud Storage Backup",IACSIT, August,2014.

[7] Weak Leakage-Resilient Client –Side deduplication of Encrypted Data in Cloud Storage" Institute for Info Comm Research,Singapore,2013

[8] Tanupriya Chaudhari , Himanshu shrivastav, Vasudha Vashisht, "A Secure Decentralized Cloud Computing Environment over Peer to Peer",IJCSMC,April,2013

[9] Mihir Bellare, Sriram keelveedhi,ThomasRistenart
,"DupLESS: Server Aided Encryption for Deduplicated storage" University of California, San Diego2013.

[10] Luna SA HSM. http://bit.ly/17CDPm1.

International Conference on, pages 617–624. IEEE, 2002.