ZKG INTERNATIONAL

ISSN: 2366-1313

ENHANCING CLOUD STORAGE SECURITY THROUGH DEDUPLICATION TECHNIQUES

#1ABDUL MASOOD,
#2YASHWANTH,
#3J.RAVI CHANDER, ***Assistant Professor,***
Department of Computer Science and Engineering,
SREE CHAITANYA INSTITUTE OF TECHNOLOGICAL SCIENCES, KARIMNAGAR, TS.

ABSTRACT: Data de-duplication is a tactic used by cloud service providers to manage unanticipated data and the challenges of cloud storage. Since the introduction of social media and the Internet of Things (IoT), there have been considerable issues related with the administration and analysis of "big data," or large amounts of information. Big data is exciting, but it poses considerable issues for cloud service providers. When faced with large amounts of data, organizations say a lot about how to manage and store it. This is especially true for providers of cloud services, as big data poses significant difficulties and opportunities. In this project, we looked at how to avoid duplicate data in cloud-based systems and highlighted the challenges that come with it. An external auditor examines the accuracy of the user's data and offers comments on its storage integrity on the cloud computing platform. Waste has lessened due to computation and communication. Before putting a user's file in the cloud, the deduplication method is used to ensure that the file already exists on the cloud computer.

Keywords: *Cloud Service Provider; Deduplication; Third Party Auditor; Data Dynamics.*

I. INTRODUCTION

Obtaining a service using cloud storage systems is ideal. The storage of data in the cloud via the Internet is referred to as "cloud storage." Many firms use cloud computing to store and retrieve data from many networked systems. The data will be completely removed from the owner's computer after it is transferred to the cloud storage server and a local copy is deleted. The security and privacy of transferred data are the most important issues in this situation. Many people lose or share data, and cloud service providers can be unreliable at times. To protect their reputation, cloud storage providers may conceal instances of data loss. To improve service-oriented accountability, the cloud server can authenticate the integrity of client data and confirm that it has not been modified. Uncertainty exists about cloud systems. Chen et al. developed the RDPC Protocol (Remote Data Possession Checking) framework. An external auditor verifies the veracity and integrity of user data stored on a cloud computer, providing cloud users with a more streamlined experience.

The number of verifications that can be done before comparing the results to the original data is not limited. This technique will save time and effort while processing and communicating. Furthermore, the suggested method uses the Merkle Hash Tree (MHT) to perform block-level data manipulation operations such as data addition, deletion, modification, and update. Deduplication is a technique for reducing data volume by removing duplicate data blocks or files. It is unnecessary to confirm that the file or material has been uploaded to the cloud. When the system detects duplicate data, it validates it and notifies the user. As a result, the store will have one identical copy. This methodology boosts storage efficiency. For deduplication, either the file or block level might be used. Non-similar files reject block-level clones of identical data blocks in the same way that file level compression does.

II. LITERATURE SURVEY

Ateniese et al.'s key goal in their research is to determine whether user data is stored on an unstable storage site. A provable data possession (PDP) paradigm may reduce the amount of file block accesses, client-server connections, and server processing, making it an appealing notion to examine. As a result, every activity requires a consistent and predetermined level of communication, which reduces the server's running costs. In their study, Shacham and Waters used the Proof of Retrievability (PoR) technique, which buried sentinels behind extra data blocks. The user selects a sentinel at random and checks the information's accuracy. Sentinels may experience damage if the storage server alters or deletes user data. As a result, the Proof of Retrievability (PoR) approach is only applicable to static and historical data storage. It is not appropriate for accessing public databases like archives, libraries, and repositories. Symmetric key cryptography is a high-performance and secure PDP alternative that does not require mass encryption. Implementing dynamic data processes makes it easier to modify and manipulate information.

The RDPC algorithm, created by Chen et al., is based on homomorphic and Merklen hash trees and allows data change. Boneh, David. Boneh, David. An efficient method for signing an item has been presented that uses bilinear mappings on elliptic curves and other techniques. This brief name is enclosed in a limit field. In November 2007, Lucca, J. et al. presented the work at the CCS '07 conference.

III. METHODOLOGY

- Cloud API
- Data Integration
- Encryption
- Hashing Algorithm
- Duplication Detection
- TPA
- Searching over encrypted data

To begin with. Implementing the dynamic RDPC protocol in cloud storage is recommended for addressing security risks, increasing efficiency, and lowering storage requirements. Rather than depending on the user to verify the data's authenticity, the proposed technique uses a third-party auditor (TPA). As a result, the individual will need to conduct fewer calculations. To hash keys, the Paillier method and the homomorphic hash function must be combined. Merkle Hash Tree (MHT) is used for procedures that need data change. The file selected by the user for storage on the cloud server is saved on their personal computer. Subsequently, the file should be separated into pieces, encrypted using the Paillier method, converted into a secure hash using a strong hash function, and finally delivered to the cloud server. The client sends a request to TPA asking if his file or data is present. TPA will then run the challenge to ensure the integrity of the file on the cloud server. As soon as TPA receives the proof from the cloud server, it will verify it and notify the user. The response will be either accurate or dishonest.

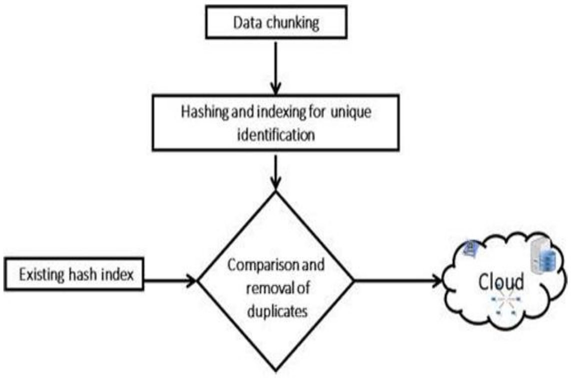

Data chunking
Hashing and indexing for unique identification
Existing hash index
Comparison and removal of duplicates
Cloud

Figure 1: Block Diagram

IV. WORKING

The new, more secure dynamic RDPC protocol includes the following key components:

Key Generation: It creates a secret key, a private key, and a public key for cryptography. Tag Production: The cloud client divides a given file F into n blocks, which are further subdivided into m sectors.

Challenge: The user is not required to initiate the challenge; the TPA will complete the task automatically.

Proof Generation: The TPA complicates things for the server. Once the calculation is completed, the proof is returned to the Trusted Third Party (TPA).

Proof Verify: After comparing evidence received from the server to evidence generated by TPA using its metadata, the application generates a response for the user.

In the absence of any damage to the data stored on the cloud server, the TPA function returns the boolean false. If the object is damaged, the function will return true.

1. Pailler Algorithm:

The Paillier algorithm is used during the key-making process. The approach used for probabilistic asymmetric encryption is known as public key cryptography.

- Key generation
- Choose any two large prime numbers arbitrarily p,q.gcd(pq, (p-1)(q-1)) =1
- Calculate value of n.
- n = pq and ƛ= lcm (p-1, q-1).
- It is possible to choose any number g at random..
- Use the modular multiplicative inverse of the existence of examination to ensure that n divides g's order.
- Data dynamics operations: Every data dynamics operation can be executed at the block level. Every time a record is updated, the verification information must be modified. As a result, modifications to operating expenses will become less frequent.

2. Hashing Algorithm:

1. Start
2. Read data owner id(udoid)
3. If (doid&amp;& amp; udoid == Null)
4. Stop
5. Read the file opening log.
6. Determine the number of blocks using the saved XML hash.
7. The user must select the block number they wish to confirm.
8. Extract the extra data for the block from the saved hash XML.
9. Using the added information, create a new root for NHT.
10. If (new root ≠ root) data modified
11. Else File not modified.
12. Stop.

3. Advanced Encryption Standard:

Advanced Encryption Standard (AES) refers to the symmetric-key block cipher technology in question. It serves as the US government's Federal Information Processing standard. The AES algorithm's key lengths of 192 and 256 are long and robust enough to protect secret data at the TOP SECRET and SECRET levels, respectively. AES requires a block size of 128 bits; however, the key length can range from 128 to 256 bits. Encryption or decryption begins with turning the receiving array of data into the State array. The output array's bytes are eventually assigned to the final State value. In AES, the number of cycles increases as the length of the key increases. For 192-bit keys, 12 rounds are used, while for 128-bit keys, it is 10. Each of the fourteen rounds will use 256-bit keys. Using the initial AES key, each round creates a unique 128-bit round key. Each AES block has four functions: SubBytes, Shift Rows, Mix Columns, or Add Round Key. To decode an AES-encrypted text, both the encryption procedure and the subkeys must be reversed. Each round of decryption includes executing the four procedures in the opposite order that they were used initially.

Unlike the AES encryption approach, the variables InvMixColumns, InvAddRoundKey, InvShiftRows, and InvSubBytes are used. Despite their close link, the decoding approach is used independently of the encryption mechanism. This is owing to the fact that it operates in the opposite direction as encryption. AES has grown in prominence in recent years as a security solution due to its interoperability with hardware and software. The goal of unbreakable AES encryption is still impossible. This can only be accomplished through intensive searching or brute force. The AES method was developed to make it more difficult for linear and differential analyses to crack codes. Because the lengths of AES keys are dynamic, exhaustive key searches may be difficult.

V. RESULTS AND DISCUSSION

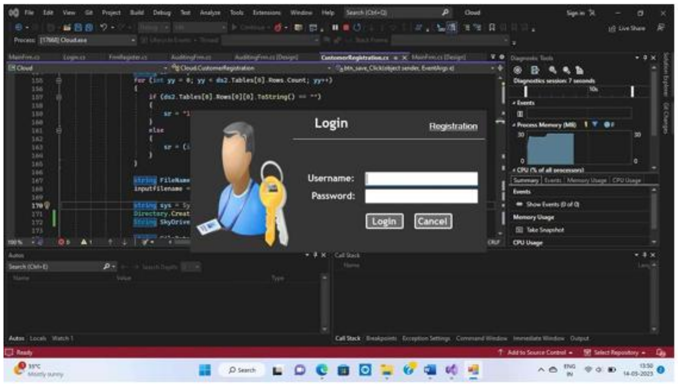
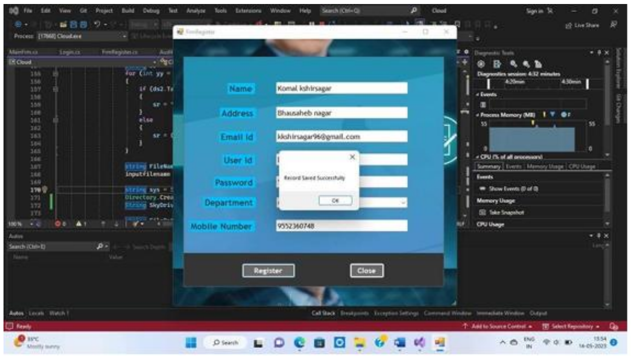


Figure 2: Login Form

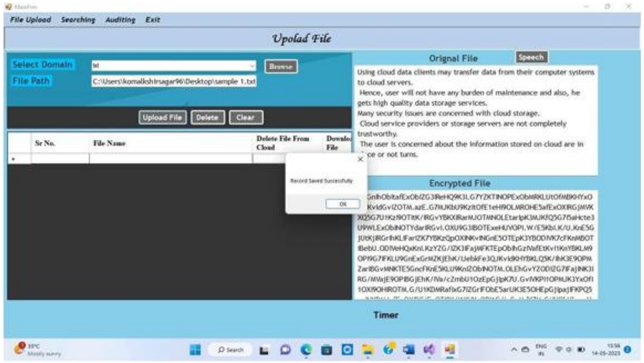


Name: Komal kshirsagar
Address: Bhausaheb nagar
Email Id: kkshirsagar96@gmail.com
User Id
Password
Department
Mobile Number: 9552360748

Record Saved Successfully

OK

Figure 3: Member Registration



Select Domain: txt
File Path: C:\Users\komalkshirsagar96\Desktop\sample 1.txt

Sr No.	File Name	Delete File From Cloud	Download File

Record Saved Successfully

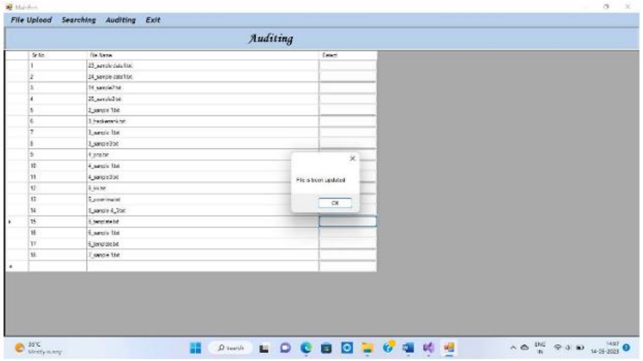
Orignal File

Using cloud data clients may transfer data from their computer systems to cloud servers.
Hence, user will not have any burden of maintenance and also, he gets high quality data storage services.
Many security issues are concerned with cloud storage.
Cloud service providers or storage servers are not completely trustworthy.
The user is concerned about the information stored on cloud are in ce or not turns.

Encrypted File

GnlhObitafExObIZG3lReHQ9K3I.G7YZKTlNOPExObMRKLUtOfMBKHYxO
KvldGvIZOTM.azE.G7MJKbU9KzltOfE1eHl9OLMROHE5afExOXIRGJMVK
XQ5G7U1Kzl9OTItK/lRGvYBKXlRarMJOTMNOLEtarlpK3MJKfQ5G7l5aHcte3
U9WLExObINOTYdarlRGvI.OXU9G3lBOTExeHUVOPI.W/E5KbLK/U.KnE5G
JUtKJlRGrlhKLlFarlZK7YBKzQpOXlNKvlNGnE5OTEpK3YBODlVK7cFKnMBOT
lBebU.ODlVeHQxKnl.KzYZG/lZK3lFaJMFKTEpObIhGzlVafEtKvl1KnYBKLM9
OPl9G7lFKLU9GnExGrMZKJEhK/UebkFe3QJKvldKHYBKLQ5K/lhK3E9OPM
ZarlBGvMNKTE5GncFKnE5KLU9KnlZObINOTM.OLEhGvYZODlZG7lFaJlNK3I
RG/MVaJE9OPlBGJEhK/lVa/cZmbU1OzEpGJlpK7U.GvlVKPl1OPMJK3YxOfI
1OXl9OHlROTM.G/U1KDMRaflxG7lZGrlFObE5arUK3E5OHEpGJlpaJlFKPQ5

Figure 4: Upload File



Sr No.	File Name	Detect
1	23_sample data1.txt	
2	24_sample data1.txt	
3	14_sample7.txt	
4	25_sample3.txt	
5	2_sample 1.txt	
6	3_hackerrank.txt	
7	3_sample 1.txt	
8	3_sample3.txt	
9	4_prac.txt	
10	4_sample 1.txt	
11	4_sample3.txt	
12	5_lo.txt	
13	5_powershell.txt	
14	5_sample 4_3.txt	
15	5_templete.txt	
16	6_sample 1.txt	
17	6_jengrade.txt	
18	7_sample 1.txt	

File is been updated

Figure 5: Auditing

VI. CONCLUSION

This method suggests an improved RDPC protocol that checks the integrity of stored data. This approach allows for block-level data alterations, such as data addition, updating, and removal. It has security procedures in place to ensure that consumer data is legitimate and confidential. Furthermore, it sends an identical cloud document, which may result in data loss and take up more space on an inactive cloud computing device. In contrast, our proposed system incorporates a deductibility mechanism that checks to see if a specific file already exists on the cloud server. The current framework is incompatible with the deduplication method. It is not necessary to save a file that is already in the cloud to make a replica. This protocol not only protects you when you connect to an untrustworthy server, but it also stops any potential attack from being modified or replaced. External monitors are contracted to facilitate cloud operations.

REFERENCES

1. M amdaqa, M., & Tahvildari, L. (2012). Cloud Computing Uncovered: A Research Landscape. H. Ali & M. Atif (Eds.), Advances in Computers Elsevier. 41–85.
2. Wang W., Zeng, G., Yao, J. (2012). Cloud-DLS: Dynamic trusted scheduling for cloud computing original research article. Expert Systems with Applications, 39(3), 2321-2329. Lin Y., Chang, P. (2011). Maintenance reliability estimation for a cloud computing network with nodes failure. Expert Systems with Applications, 38(11), 14185-14189.
3. Chen, L., Zhou, S., Huang, X., Xu, L. (2013). Data dynamics for remote data possession checking in cloud storage. Computers and Electrical Engineering, 39, 2413-2424.
4. Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou. A Hybrid Cloud Approach for Secure Authorized Deduplication. IEEE Transactions On Parallel And Distributed System Vol:Pp No:99 (2014).
5. Ateniese, G., Burns, R., Curtmola, R., Herring, J., Kissner, L., Peterson, Z., et al. (2007) Provable data possession at untrusted stores. In ACM CCS 2007, ACM, 598–609
6. Shacham, H., & Waters, B. (2008). Compact proofs of retriev ability. ASIACRYPT 2008 (Vol. 5350, pp. 90–107). Berlin/ Heidelberg: Springer.
7. Ateniese, G., Pietro, R. D., Mancini, L. V., &Tsudik, G. (2008). Scalable and efficient provable data possession. In SecureComm'08.
8. Erway, C., Kupcu, A., Papamanthou, C., & Tamassia, R. (2009). Dynamic provable data possession. In ACM CCS'09 (pp. 213–222).
9. Wang, Q., Wang, C., Ren, K., Lou, W., & Li, J. (2012). Enabling public auditability and data dynamics for storage security in cloud computing. IEEE Transactions on Parallel and Distributed Systems, 22(5), 847 859.
10. Yong Yu, Jianbing Ni, Man Ho Au, Hongyu Liu, Hua Wang, Chunxiang Xu. Improved security of a dynamic remote data possession checking protocol for cloud storage. Expert Systems with Applications 41 (2014).