Abstract

This paper discusses the current quandary faced in the cloud computing with regard to preserving the privacy in sharing the data and the deduplication of the data. These accommodations are provided from data centers which are located throughout the world. Data deduplication is a technique for eliminating duplicate copies of data, and has been widely used in cloud storage to reduce storage space and upload bandwidth. However, there is only one copy for each file stored in cloud even if such a file is owned by a huge number of users. Aiming to address the above security challenges, this paper makes the first attempt to formalize the notion of distributed reliable deduplication system. We propose new distributed deduplication systems with higher reliability in which the data chunks are distributed across multiple cloud servers. Security analysis demonstrates that our deduplication systems are secure in terms of the definitions specified in the proposed security model. As a proof of concept, we implement the proposed systems and demonstrate that the incurred overhead is very limited in realistic environments. The proposed system provides a solution for preserving the data in cloud with the avail of encryption protocol. Three modules are presented here namely data owner, Third Party Administrator, and retailer.
I. INTRODUCTION

Presently a day there is development in data. With limitless storage room give by cloud specialist co-op. Clients will get to data as indicated by their necessities and most clients get to same data over and over, the cost of calculation, application facilitating, content stockpiling and conveyance is diminished altogether. The cloud making it workable for you to get to your data from anyplace whenever. Cloud gives advantages, for example, adaptability, recuperation, programming refreshes consequently, pay-per-utilize model and cost reduction.[3] While a conventional PC setup obliges you to be in an indistinguishable area from your information stockpiling gadget. The cloud expels the requirement for you to be in an indistinguishable physical area from the equipment that stores your information.

Information deduplication is a system for diminishing the measure of storage room an association needs to spare its information. In many associations, the capacity frameworks contain copy duplicates of many bits of information. For instance, a similar record may spare in a wide range of spots by various clients, or at least two documents that aren't indistinguishable may at present incorporate a significant part of similar information. To make information administration adaptable deduplication we are utilize Encryption for secure deduplication services.[1]

There are two sorts of deduplication as far as the size: (i) file-level deduplication, which expels redundancies between various files and (ii) block level deduplication, which find and evacuate duplication between information pieces. The file can be isolated into littler fixed-size or variable-estimate squares. Utilizing fixed size pieces simplifies the calculations of square limits, while utilizing variable-estimate squares. A method which has been proposed to meet these two conflicting necessities is Tag era and AES Scheme whereby the encryption key is generally the aftereffect of the hash of the information portion; it tragically experiences different surely understood shortcomings. The secrecy issue can be taken care of by scrambling delicate information before outsourcing to remote servers.[13]

Alongside low proprietorship expenses and adaptability, clients require the insurance of their information and secrecy ensures through encryption.[9]

In this paper, we address the a for specified protection issue to propose a mutual expert to the records which Deduplicated based security saving validation for the cloud information stockpiling, which acknowledge verification without trading off a client’s private data. The primary commitment is as per the following.

Figure 1: Cloud Computing Platform

Ravi Rathod, Akshada Chorage, Bhagyashri Landge, Kadambari Londhe, Shilpa Waghmare :: A Novel Approach for Secure Data Deduplication Using Shared Authority in Cloud Storage
1) Identifies another security challenge in distributed storage, and address an unobtrusive protection issue, in which the tested demand itself can't uncover the client's protection regardless of whether or not it can get the getting to expert.

2) Proposes a confirmation convention to upgrade a client's get to ask for related security, and the mutual get to specialist is accomplished by unsecured get to ask for coordinating component.

3) Apply figure content arrangement credit based get to control to understands that a client can dependably get to its own particular information gelds, and receive the intermediary re-encryption to give temp approved information sharing among various users.[10]

II. RELATED WORK

- The tested get to demand itself may uncover the client's security regardless of whether or not it can acquire the information get to authorizations. Hong Liu, Huansheng Ning,[10] they propose a common expert based security saving validation convention (SAPA) to address above protection issue for distributed storage. In the SAPA, 1) shared get to expert is accomplished by unknown get to ask for coordinating component with security and protection contemplations (e.g., confirmation, information namelessness, client protection, and forward security); 2) characteristic based get to control is embraced to understand that the client can just get to its own particular information fields; 3) intermediary re-encryption is connected by the cloud server to give information sharing among the different clients. In the interim, all inclusive compensability (UC) models is set up to demonstrate that the SAPA hypothetically has the outline rightness.

- Boyang Wang, Baochun Li,[6] they propose the primary protection saving system that permits open inspecting on shared information put away in the cloud. Specifically, we abuse ring marks to figure the confirmation data expected to review the respectability of shared information. With our system, the personality of the endorser on each square in shared information is kept private from an outsider evaluator (TPA), who is as yet ready to freely confirm the respectability of shared information without recovering the whole document. Our exploratory outcomes exhibit the viability and effectiveness of our proposed component when examining shared information.

- Ayad F. Barsoum and M. Anwar Hasan,[7] They propose a guide based provable multicopy dynamic information ownership (MB-PMDDP) plot that has the accompanying components: 1) it gives a proof to the clients that the CSP is not tricking by putting away less duplicates; 2) it underpins outsourcing of element information, i.e., it bolsters piece level operations, for example, square adjustment, inclusion, erasure, and annex; and 3) it permits approved clients to consistently get to the record duplicates put away by the CSP. We give a near examination of the proposed MB-PMDDP plot with a reference demonstrate got by developing existing provable ownership of element single-duplicate plans. The hypothetical investigation is approved through test comes about on a business cloud stage. Furthermore, we demonstrate the
security against conspiring servers, and examine how to distinguish defiled duplicates by somewhat altering the proposed plot.

- T. Sivashakthi1, Dr. N Prabakaran, [3] from the previous couple of years, there has been a quick advance in cloud, with the expanding quantities of organizations turning number of organizations depending on utilize assets in the cloud, there is a need for ensuring the information of different clients utilizing concentrated asset. Distributed storage administrations stay away from the cost stockpiling administrations keeps away from the cost costly on programming, work force keeps up and gives better execution less capacity cost and adaptability, cloud benefits through web which increment their introduction to capacity security vulnerabilities however security is one of the significant downsides that averting huge associations to go into distributed computing condition. This work reviewed on a few stockpiling systems and these points of interest and its disadvantages.

- Pasquale PuzioSecludIT and EURECOM Sophia-Antipolis, [9] they propose ClouDedup, a protected and proficient stockpiling administration which guarantees piece level deduplication and information privacy in the meantime. Albeit in light of united encryption, ClouDedup stays secure on account of the meaning of a segment those executes an extra encryption operation and a get to control system. Moreover, as the necessity for deduplication at square level raises an issue as for key administration, we propose incorporating another part so as to actualize the key administration for each piece together with the real deduplication operation. We demonstrate that the overhead presented by these new segments is negligible and does not affect the general stockpiling and computational expenses.

### III. Proposed Methodology

- The term deduplication characterizes that the way toward disposing of the copy duplicates or same duplicates in the distributed storage. Also, the mutual specialist characterizes to give the security to the document which are been transferred in the cloud
- The first process the client needs to login in the cloud and on the off chance that he has effectively enlisted his record. At that point straightforwardly login in the sytem. And after the enrollment of the client all the data is been spared in detail in the database.
- Data proprietor and client is the person who will transfer the document on cloud, as the record will be transferred on cloud the primary concern it will do is to check the copy duplicates of the record which has been transferred by the client.
- The deduplication of the record will be done piece savvy deduplication each word in the document will be crossed checked and after that the following procedure if any copy duplicate is found that document will dispose of and if not the document will be prepared to transfer on cloud.
- The document which is to be transferred will be in scrambled configuration so that the other client won't have the capacity to get to the information from the record.
The second most essential piece of the venture is shared specialist as the record will be transferred on cloud if the other client needs to get to the clients document he needs to send a demand to the cloud separate segment and afterward the third gathering will request that the information proprietor acknowledge. Ask for on the off chance that he needs to permit the other client to get to his information if not then the authorization will be denied.

if the information proprietor needs to permit the other client to get to the information if will acknowledge the demand from the outsider and the following they will be changed over in unscrambled design so that the other client will have the capacity to get to the information from the document.

And as the document is in decoded arrange the client can download the record from distributed storage.

3.1 Proposed System Architecture

3.2 System Module
1. Client Registration:
In this module a client/proprietor needs to transfer its records in a cloud server, he/she ought to enroll first. At that point just he/she can have the capacity to do it. For that he requires to fill the points of interest in the enrollment frame. These subtle elements are filled in a database.

2. Client Authenticate:
In this module, any of the previously mentioned individual need to verify, they ought to perceived by giving their email-id and secret word

3. Use Registration:
In this module if a use needs to get to the information which is put away in a cloud, he/she ought to enlist their subtle elements first. These points of interest are put away in a Database.
4. **Use Authenticate:**
In the event that the utilizer is an approved utilizer, he/she can download the record by using document id which has been put away by information proprietor when it was transferring a document/information.

5. **Information Deduplication:**
Check or Deduplicate record/information utilizing Block level Deduplication and label Generation calculation for piece level deduplication procedure. We will dispose of the Duplicate document. Unique record will transfer on cloud server.

6. **S-CSP:**
An S-CSP is a safe cloud specialist organization and CSP who handles cloud servers (CSs) and offers paid storage room on its foundation to store the proprietor's documents.

7. **Encryption and Decryption:**
Here we are using this aes_encrypt and aes_decrypt for encryption and decoding. The record we have transferred which must be in scrambled frame and decode it.

8. **Record Upload:**
In this module Owner transfers the file (along with Meta information) into database, with the profit of this metadata and its substance, the allegations utilizer needs to download the record. The transferred record was in scrambled shape; just enrolled utilizer can decode it.

9. **Document Download:**
The approved clients can download the document from cloud database.

10. **TTP (trusted outsider) validates:**
In this module TTP has screens the information proprietors record by confirming the information proprietor's document and put away the document in a database. Additionally TTP checks the CSP (cloud convenience supplier), and discover whether the CSP is permitted one or not.

11. **Focal points:**
• Improve the unwavering quality of information
• Achieving the privacy of the clients' outsourced information.
• Unique highlight of the proposition is that information trustworthiness, and also label consistency, can be accomplished.
• Here we proposed the secured framework and information proprietor can choose whether the client can get to the framework or not.

### 3.3 Algorithm
#### 3.3.1 RSA algorithm

**Key Generation**

1. Choosing two very large prime numbers p and q.
2. Compute their system modulus, \( n = p \times q \) and the ‘totient’ function \( \phi(n) = (p - 1)(q - 1) \). Note that the factors \( p \) and \( q \) remain secret and \( n \) is public.
3. Select the encryption key \( e \) at random, so that \( \gcd(e, \phi(n)) = 1 \), where \( 1 < e < \phi(n) \).
4. Solve the following equation to find the decryption key \( d \):
(5) Publish the public encryption key: \( PU = \{e, n\} \), which is known to everyone.
(6) Keep secret or private the decryption key: \( PR = \{d, n\} \), which is known only to the person who has to decrypt or sign the message.

**Data Encryption**

(1) Input the plaintext or message \( M \), where \( 0 \leq M \leq n \).
(2) Obtain the public key of recipient, \( PU = \{e, n\} \).
(3) Compute the cipher \( C \), using the following equation: \( C = M^e \mod n \)

**Data Decryption**

(1) Input the cipher text \( C \).
(2) Use their private key, \( PR = \{d, n\} \).
(3) Compute the message \( M \), using the following equation: \( M = C^d \mod n \)

### 3.3.2 SHA algorithm

**Step 1:** Append Padding Bits.... Message is “padded” with a 1 and as many 0’s as necessary to bring the message length to 64 bits fewer than an even multiple of 512.

**Step 2:** Append Length....

64 bits are appended to the end of the padded message. These bits hold the binary format of 64 bits indicating the length of the original message.

**Step 3:** Prepare Processing Functions....

SHA1 requires 80 processing functions defined as:

- \( f(t;B,C,D) = (B \text{ AND } C) \text{ OR } ((\text{NOT } B) \text{ AND } D) \quad (0 \leq t \leq 19) \)
- \( f(t;B,C,D) = B \text{ XOR } C \text{ XOR } D \quad (20 \leq t \leq 39) \)
- \( f(t;B,C,D) = (B \text{ AND } C) \text{ OR } (B \text{ AND } D) \text{ OR } (C \text{ AND } D) \quad (40 \leq t \leq 59) \)
- \( f(t;B,C,D) = B \text{ XOR } C \text{ XOR } D \quad (60 \leq t \leq 79) \)

**Step 4:** Prepare Processing Constants....

SHA1 requires 80 processing constant words defined as:

- \( K(t) = 0x5A827999 \quad (0 \leq t \leq 19) \)
- \( K(t) = 0x6ED9EBA1 \quad (20 \leq t \leq 39) \)
- \( K(t) = 0x8F1BBCDC \quad (40 \leq t \leq 59) \)
- \( K(t) = 0xCA62C1D6 \quad (60 \leq t \leq 79) \)

**Step 5:** Initialize Buffers....

SHA1 requires 160 bits or 5 buffers of words (32 bits):

- \( H0 = 0x67452301 \)
- \( H1 = 0xEFCDAB89 \)
- \( H2 = 0x98BADCFE \)
- \( H3 = 0x10325476 \)
- \( H4 = 0xC3D2E1F0 \)

**Step 6:** Processing Message in 512-bit blocks (L blocks in total message)....

This is the main task of SHA1 algorithm which loops through the padded and appended message in 512-bit blocks.

Input and predefined functions:

- \( M[1, 2, ..., L] \): Blocks of the padded and appended message
- \( f(0;B,C,D), f(1;B,C,D), ..., f(79;B,C,D) \): 80 Processing Functions
- \( K(0), K(1), ..., K(79) \): 80 Processing Constant Words

\( H0, H1, H2, H3, H4, H5 \): 5 Word buffers with initial values
4. IMPLEMENTATION AND EVALUATION

Implementation is the stage of the project when the theoretical design is turned out into a working system. Thus it can be considered to be the most critical stage in achieving a successful new system and in giving the user, confidence that the new system will work and be effective. The implementation stage involves careful planning, investigation of the existing system and its constraints on implementation, designing of methods to achieve changeover and evaluation of changeover methods.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Proposed Approach</th>
<th>Two-Phase Deduplication</th>
<th>Multi-Layered Cryptosystem</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space</td>
<td>Partial Deduplication in filelevel means more space saving</td>
<td>Sample file level deduplication is provided</td>
<td>Not implemented</td>
</tr>
<tr>
<td>Security</td>
<td>Authorized deduplication</td>
<td>Inter-user deduplication is less secure</td>
<td>Provide more privacy to user who owns unpopular file while privacy degraded for users with popular files.</td>
</tr>
<tr>
<td>Bandwidth</td>
<td>If file will get duplicate then it will not uploaded</td>
<td>In intra user bandwidth saving ,in inter user more bandwidth saving</td>
<td>Not implemented</td>
</tr>
<tr>
<td>Computational Overhead</td>
<td>Using RSA algorithm for encryption and decryption</td>
<td>Not implemented</td>
<td>Overhead is caused by convergent and symmetric encryption</td>
</tr>
</tbody>
</table>

IV. EXPERIMENTAL RESULTS

In this section we discus one example to show that how our work is implemented. After that we compare the evaluation parameter of our proposed approach with existing approach. Consider a college has a private cloud. Consider a user of that college who has uploaded right to the public cloud wants to upload a word file. First the file gets divided into pages and after that a hash key is generated per page. That hash key get stored on to the private cloud of the college. After the encryption of the file is done using AES algorithm on to the private cloud and then the file get stored to the public cloud. If another user want to upload the same document with some minor changes ,consider 2 or 3 pages then hash key is generated for only that 2 or 3 pages. The hash key is generated from the content of that file. And it will get shown on the partial duplicate document section.
A Novel Approach for Secure Data Deduplication using Shared Authority in Cloud Storage

Figure 3: Homepage of the system

Figure 4: Registration page of the system

Figure 5: Registration completed

Ravi Rathod, Akshada Chorage, Bhagyashri Landge, Kadambari Londhe, Shilpa Waghmare :: A Novel Approach for Secure Data Deduplication Using Shared Authority in Cloud Storage
A Novel Approach for Secure Data Deduplication using Shared Authority in Cloud Storage

**Figure 6:** User can login after the completion of the registration.

**Figure 7:** As the user has already created the account user can login in the system.

**Figure 8:** User can upload the file.
Figure 9: File is been converted into encryption format to provide the security to the file.

Figure 10: To check the status of the file after been uploaded

Figure 11: To check the record of the files which are been uploaded
V. CONCLUSION

We proposed the distributed deduplication systems to improve the reliability of data while achieving the confidentiality of the users and also shared authority outsourced data with an encryption mechanism. Four constructions were proposed to support file-level and block-level data deduplication. The security of tag consistency and integrity were achieved. We
implemented our deduplication systems using the Ramp secret sharing scheme or tag generation scheme and demonstrated that it incurs small encoding/decoding overhead compared to the network transmission overhead in regular upload/download operations. In this work, we have identified a new privacy challenge during data accessing in the cloud computing to achieve privacy-preserving access authority sharing for deduplicated files. Authentication is established to guarantee data confidentiality and data integrity. It indicates that the proposed scheme is possibly applied for enhanced privacy preservation in cloud applications.

VI. REFERENCES

[1] Jin Li, Xiaofeng Chen, Xinyi Huang, Shaohua Tang and Yang Xiang Senior Member, IEEE and Mohammad Mehdiz Hassan Member, IEEE and AbdulhameedAlelaiwi Member, IEEE “Secure Distributed Deduplication Systems with Improved Reliability” IEEE Transactions on Computers, 2015


[10] Hong Liu, Student Member, IEEE, HuanshengNing, Senior Member, ieee, QingxuXiong, Member, IEEE,and Laurence T. Yang, Member, IEEE “Shared Authority Based Privacy-preserving Authentication Protocol in Cloud Computing” IEEE transactions on parallel and distributed systems year 2014


