RESEARCH ARTICLE OPEN ACCESS

# SDDMCSS: Secure and Data De-duplicated Multi-Cloud Storage System

Priyanka Sunil Rokade[1] , Biplab Kumar Sarkar[2] , Prof. Dr. Bej Raj Singh Patel[3]
[1]Student, Department of Computer Engineering, PVPIT, Pune India
[2, 3] Professor, Department of Computer Engineering, GEH-TCS, Pune India

----------**********----------

## Abstract:

Normally when user uploads any file that file is stored on single cloud service provider. From a customer's point of view relying upon a solo Service Provider for his outsourced data is not very promising. In addition, providing better privacy as well as ensuring data availability, can be achieve by dividing the user's data block into data pieces and distributing them among the available Service Providers in such a way that no less than a threshold number of Service Providers can take part in successful retrieval of the whole data block. In the proposed system a secured cost-effective multi-cloud storage (SCMCS) model in cloud computing which holds an economical distribution of data among the available Service Providers in the market, to provide customers with data availability as well as secure storage. Along with file storage on multiple clouds, here invention focusses on 'Data Deduplication' technique also. Deduplication is a method of removing duplicate files or data from the storage, and is used in cloud computing environment to minimize storage space as well as network bandwidth. In data deduplication single copy of file is stored in cloud environment, which is owned by a large number of data users which may leads to reduced reliability, availability and security. Along with deduplication, there is a major issue of security for sensitive data, when the data has been outsourced on the cloud. By addressing security challenges, we attempt to formalize the secure distributed deduplication system. Here we are proposing new distributed data deduplication system, which achieves more confidentiality, integrity and reliability, as compare to the traditional deduplication system. In the proposed system single data chunk or file is divided and distributed on the multiple cloud storage servers, in such a way that the single part of file is unpredictable. This is done by introducing a Shamir secret sharing scheme and hashing algorithms in distributed cloud storage system, without using traditional ways of encryption-decryption scheme. Also proposed system will be focusing on the recovery and reconstruction of corrupted data or failed storage site without using traditional backup or recovery methods as like RAID array method.

*Keywords*— **SDDMCSS, data deduplication, cloud storage, reliability, availability and RAID array.**

----------**********----------

## I. INTRODUCTION

A number of deduplication systems have been proposed based on various deduplication strategies such as client-side or server-side deduplications, file-level or block-level deduplications. According to the data granularity, deduplication strategies can be categorized into two main categories: file-level deduplication and block-level deduplication, which is nowadays the most common strategy. In block-based deduplication, the block size can either be fixed or variable. Another categorization criteria is the location at which deduplication is performed: if data are deduplicated at the client, then it is called source-based deduplication, otherwise target-based. In source-based deduplication, the client first hashes each data segment he wishes to upload and sends these results to the storage provider to check whether such data are already stored: thus only non-deduplicated data segments will be actually uploaded by the user. While deduplication at the client-side can achieve bandwidth savings, it unfortunately can make the system vulnerable to side-channel attacks whereby attackers can immediately discover whether a certain data is stored or not. On the other hand, by deduplicating data at the storage provider, the system is protected against side-channel attacks but such solution does not decrease the communication overhead.

**'Secure and data de-duplicated multi-cloud storage system (SDDMCSS)'** which describes multiple functionalities. 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. As a result, deduplication system improves storage utilization while reducing reliability. Furthermore, the challenge of privacy for sensitive data also arises when they are outsourced by users to cloud. Aiming to address the above security challenges, it makes the first attempt to formalize the notion of distributed reliable deduplication system. In proposed system data chunks are distributed across multiple cloud servers. The security requirements of data confidentiality and tag consistency are also achieved by introducing a deterministic secret sharing scheme in distributed storage systems, instead of using convergent encryption as in previous deduplication systems. A number of deduplication systems have been proposed based on various deduplication strategies such as client-side or server-side deduplications, file-level or block-level deduplications. Especially, with the advent of cloud storage, data deduplication techniques become more attractive and critical for the management of ever-increasing volumes of data in cloud storage services which motivates enterprises and organizations to outsource data storage to third-party cloud providers, as evidenced by many real-life case studies. There are two types of deduplication in terms of the size:

- File level De-duplication: which discovers redundancies between different files and removes these redundancies to reduce capacity demands.
- Block level De-duplication : which discovers redundancies between different files and removes these redundancies to reduce capacity demands provides better deduplication efficiency.

Data reliability is actually a very critical issue in a deduplication storage system because there is only one copy for each file stored in the server shared by all the owners. If such a shared file/chunk was lost, a disproportionately large amount of data becomes inaccessible because of the unavailability of all the files that share this file/chunk. If the value of a chunk were measured in terms of the amount of file data that would be lost in case of losing a single chunk, then the amount of user data lost when a chunk in the storage system is corrupted grows with the number of the commonality of the chunk. Thus, how to guarantee high data reliability in deduplication system is a critical problem. Proposed system defines data deduplication along with that it stress copy of every data in such a way that single part of data item is unpredictable to user so that user can not generate original data from this single part of data. Also these small shares are stored on multiple cloud service provider storages to improve more security. Here even if whole storage site is corrupted or hacked by unauthorized user then also user data is safe. Data reconstruction is possible from other storage site data and database which avoids use of extra backup methods.

## II. Related Work

Ashutosh Saxena, Nitin Singh Chauhan, Sravan Kumar Rondla have defined, “System and method for verifying integrity of cloud data using unconnected trusted device” US 9641617 B2

The present invention provides a method and system for verifying integrity of cloud data using unconnected trusted device. The method involves requesting encrypted data though a terminal from a metadata offsite location on a cloud storage then entering encrypted data into an unconnected trusted device thereafter obtaining sentinel data from one or more predefined sentinel locations in encrypted data then requesting original data from the cloud storage through the terminal from the unconnected trusted device thereafter comparing sentinel data and original data for integrity and finally displaying the results.[1].

Bogdan Nicolae have defined, “Cost-effective IAAS (infrastructure-as-a-service) cloud storage based on adaptive virtual disks (AVD)” US 9442669 B2

There are provided a system, a method and a computer program product for operating a cloud computing storage system. The cloud computing storage system allocates and manages virtual disks. A virtual disk provides a logical data storage. The cloud computing storage system divides data stored in the virtual disks into chunks and allocates the chunks to physical data storage devices. The cloud computing storage system monitors I/O access patterns and user requests to change data storage capacities and throughputs of the virtual disks in real time. The cloud computing storage system dynamically reconfigures an allocation of the chunks to the physical data storage devices.[2].

Rahul V. Auradkar, Roy Peter D'Souza have defined, “Secure and private backup storage and processing for trusted computing and data services” US 20100318812 A1

A digital escrow pattern is provided for backup data services including searchable encryption techniques for backup data, such as synthetic full backup data, stored at remote site or in a cloud service, distributing trust across multiple entities to avoid a single point of data compromise. In one embodiment, an operational synthetic full is maintained with encrypted data as a data service in a cryptographically secure manner that addresses integrity and privacy requirements for external or remote storage of potentially sensitive data. The storage techniques supported include backup, data protection, disaster recovery, and analytics on second copies of primary device data. Some examples of cost-effective cryptographic techniques that can be applied to facilitate establishing a high level of trust over security and privacy of backup data include, but are not limited to, size-preserving encryption, searchable-encryption, or Proof of Application, blind fingerprints, Proof of Retrievability, and others.[3]

Jonathan M. Barney, Cataldo Mega, Edmond Plattier, Daniel Suski have defined, “Method and system for managing security in a computing environment” US 9560019 B2

A method and system for managing data security in a computing environment. A processor at the gateway server receives, from a user device, at least one message. Each

message requests that an encryption key be downloaded to the user device. The gateway server interfaces between the user device and a cloud that includes interconnected computing systems external to the user device. In response to the received at least one message, the processor generates at least one unique encryption key for each message and sends the at least one generated encryption key to the user device, but does not store any of the generated encryption keys in the cloud. For each encryption key having been sent to the user device, the processor receives each encryption key returned from the user device. For each encryption key received from the user device, the processor stores each received encryption key in the cloud.[4].

Jens-Matthias Bohli, Ghassan KARAME have defined, "A method for storing of data within a cloud storage and a cloud storage system" WO 2016026537 A1
For providing an alternative secure cloud storing of data, additionally allowing fair billing of storing data within a cloud storage, a method for storing of data within a cloud storage is claimed, wherein data of a user is stored within the cloud storage upon a request by the user. The method is characterized in that the data is encrypted, the request is directed to a managing means and - before an uploading of the encrypted data to the cloud storage - the managing means performs a deduplication on the encrypted data, so that uploading of the data is only performed, if the data is not yet stored within the cloud storage. Further, an according cloud storage system is claimed, preferably for carrying out the above mentioned method.[5].

David Lanc, Lu Fan, Lachlan MACKINNON, Bill BUCHANAN have defined, "Resilient secret sharing cloud based architecture for data vault" US 20170005797 A1
A method of securely storing data including: providing, within a secure data storage system, a plurality of secret sharing methods for selection and identifying a striping policy for storage of the data, in accordance with input preferences. The data can be split into N secret shares according to a secret sharing method, the selection being determined by the striping policy, wherein a threshold number, T, of such shares is sufficient to recover the data, where T is less than N, generating metadata associated with the data, the metadata identifying the selected secret sharing method and storing the metadata within the secure data storage system and writing the secret shares to storage that includes storage outside the secure data storage system, such that, when at least T shares are retrieved, the metadata can be recalled to identify the selected secret sharing method for recovery of the data.[6].

Ashutosh Saxena, Nitin Singh Chauhan, Sravan Kumar Rondla have defined, "System and method for verifying integrity of cloud data using unconnected trusted device" US 9641617 B2
The present invention provides a method and system for verifying integrity of cloud data using unconnected trusted device. The method involves requesting encrypted data though a terminal from a metadata offsite location on a cloud storage then entering encrypted data into an unconnected trusted device thereafter obtaining sentinel data from one or more predefined sentinel locations in encrypted data then requesting original data from the cloud storage through the terminal from the unconnected trusted device thereafter comparing sentinel data and original data for integrity and finally displaying the results.[7].

A Secured Cost Effective Multi -Cloud Storage in Cloud Computing
Prof.V.N.Dhawas, Pranali Juikar ,Neha Patekar ,Neha Lendghar,Sushant Vartak
The term "Cloud" is analogical to "Internet". Cloud computing is Internet based computing where virtual shared servers provide software, infrastructure, platform, devices and other resources and hosting to customers on a pay -as-you
-use basis. Cloud data storage redefines the security issues targeted on customer's outsourced data. From a customer's point of view relying upon a solo Service Provider for his outsourced data is not very promising. In addition, providing better privacy as well as ensuring data availability, can be achieve by dividing the user's data block into data pieces and distributing them among the available Service Providers in such a way that no less than a threshold number of Service Providers can take part in successful retrieval of the whole data block.In this paper,we propose a secured cost –effective multi-cloud storage(SCMCS)model in cloud computing which holds an economical distribution of data among the available Service Providers in the market, to provide customers with data availability as well as secure storage.[8].

Secure Distributed Deduplication Systems with Improved Reliability:
The paper provides confidentiality, integrity and reliability using RSSS and tag consistency. They focused on achieving the data reliability. Here they proposed distributed deduplication system along with secret sharing scheme instead of using convergent encryption scheme. Jin Li have used Ramp secret sharing scheme for the data splitting technique. Tag consistency used for integrity purpose and tag generation is done by the end user. This scheme gives kind of workload to the end user.[9].

Secure deduplication with efficient and reliable convergent key management:
The basic idea of convergent encryption (CE) uses Dekey, used for encryption decryption. Use of secret sharing for security of key. Dekey constructs secret share on plain text and distributes across multiple cloud service providers. Dekey can be used by multiple users instead of different key for different users. Here multiple user shares the same block, due to which storage space is minimized. They focused only on the confidentiality of data. In this work cipher text can be easily duplicated.[10].

In DupLESS, clients encrypt under message-based keys obtained from a key-server.
Group of clients encrypt data with the help of key server which is separate from storage server. Clients have to authenticate themselves to the key server. But the point of failure is key server, unless key server is secure whole system is secure. DupLESS technique achieves strong confidentiality.[11].

Message-locked encryption and secure de-duplication:
Message-Locked Encryption (MLE), where the key under which encryption and Decryption are performed, derived from the message. Message is mapped to the key for encryption and decryption technique. Here symmetric encryption scheme has been used by clients for secure deduplication. This scheme provides both privacy and integrity of data. But this scheme works fine with single client not for multiple clients.[12].

A secure data deduplication scheme for cloud storage:
An encryption scheme that guarantees semantic security for unpopular data and provides weaker security, better storage and bandwidth benefits for popular data and unpopular data. Popular data are normally not sensitive hence the traditional encryption mechanism is performed. For unpopular data they have provided semantic security and multi-layered cryptosystem for supporting deduplication. So they provides security to the outsourced data in the cloud deduplication system.[13].

## III. SECURE AND DATA DE-DUPLICATED MULTI-CLOUD STORAGE SYSTEM(SDDMCSS)

'Secure and data de-duplicated multi-cloud storage system (SDDMCSS)', which performs data deduplication at smallest level to save storage space and bandwidth along with providing security, integrity, confidentiality and reliability to user data. It uses multiple cloud storage servers to store the copy of user data. It stores single copy of user data to avoid data redundancy. Once user uploads a file it checked for various kind of data deduplication techniques like file level, block level, fixed size block level, variable size block level etc. using different hashing algorithm scheme. At file level two different hash algorithms are used and for block level also two different hash algorithms are used to check data deduplication. This scheme achieves more security than existing data deduplication scheme. After data deduplication check sharing scheme is used to store data on multiple cloud storage servers. These multiple cloud storage servers' stores share of data files.

Single share of data file is unpredictable to user, so that any unauthorized person unable to understand the original data. Also various types of attacks can be avoided like chosen distributed attack, collision attack, insider attack etc. Due to use of distributed cloud storage server's availability and reliability of data has been increased. Also due to use of secret sharing scheme corrupted data or storage site can be easily re-established with the help of database. Thus use of multiple cloud storage server and hashing scheme enhances security of de-duplicated data.

## IV. IMPLEMENTATION DETAILS

### A. *'Secure and Data de-duplicated multi-cloud storage system(SDMCSS)'*

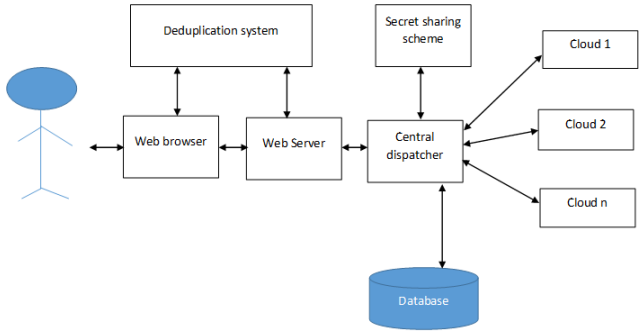


Fig. 1.1 'Secure and data de-duplicated multi-cloud storage system(SDDMCSS)'

Explanation: Fig. 1.1 shows the complete block diagram of 'Secure and data de-duplicated multi-cloud storage system (SDDMCSS)'. It comprises of web browser, web server, deduplication system, central dispatcher, secret sharing scheme, multiple cloud service providers and database. Proposed system focused on the file level and block level, client side and server side deduplications. As shown in Fig. 1.1. When user wants to upload the file on the cloud storage at that time user requests to the web browser for uploading the file. Only approved user can upload or download the file through web server. To check the authorized user we are calculating hash value of the data at file and block level using two different hashing algorithms (Rabin fingerprint and MD5), those hash values are shared with users and servers as a proof of ownership. By matching these hash values we are easily detect the original user and attacker. Process to be followed in upload and download operation is shown in Fig.1.1 When file is uploaded, the original file splits into fixed size blocks. According to the file size the number of fixed size blocks are created. After this process deduplication check occurs. Data storage server contains all the uploaded files in the form of secret share of that files along with hash value. With the help of share and recover, algorithms of secret share encryption and decryption takes place. Database stores metadata of the files and users. The distributed Deduplication system stores data on the distributed multiple cloud service provider storages. It distributes the data across multiple cloud storage servers more reliably. Our new construction uses the technique of secret splitting to split data item into multiple secret shares, and distributes these shares across multiple distributed storage servers. For that purpose any secret splitting scheme can be used, which divides and distribute

different shares across various servers and combine those shares to recover the original data. Due to the distributed feature we can achieve high availability and fault tolerance. On the distributed sites central dispatcher dispatches secret share and their hash values. File and fine grained block level deduplication check occurs.

Following steps to be followed for whole operation.

Initializes a secret share scheme that is SS (Secret Share) = (Share,Recover). This share and recover used to split the secret and recover the data from the splitted secrets.

_ File upload: Firstly check file level deduplication with database. Consider two

Cases :

– If a duplicate is found: then user will get the message "Your file has been uploaded", and user's pointer to that particular file is created by the system.

– If no duplicate is found: In this case, block level deduplication check comes in picture. Same process is done here as like file level deduplication except here we have created block pointers with user's information.

_ File download: In this step when user downloads a file central dispatcher collects all the shares from the distributed sites then checks integrity, recover the blocks to the main file and then provided to end user. Different shares are stored on the different cloud service providers.

Data modification Operations like create, read, write and update takes place in a secure way on the shared data blocks such that no user will be affected by the other users modification on the stored data, if both users pointed to the same file. Modified data will be shown to the respective users only. Collusion attacks handled properly. Here we can recover the corrupt data or failure site, with the help of Secret share algorithm. Also one of the more important feature of the proposed system is we can reconstruct the failure site, with the help of database and secret share algorithm without using traditional recovery ways like RAID etc.

*B. Diagram of Secure and data de-duplicated multi-cloud storage system (SDDMCSS)*

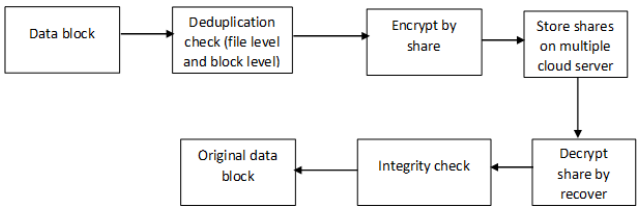


Fig. 1.2 Diagram of Secure and data de-duplicated multi-cloud storage system (SDDMCSS)

.

Explanation: Fig. 1.2 is a diagram of 'Secure and data de-duplicated multi-cloud storage system (SDDMCSS)'. It comprises of data block, Deduplication check (file level and block level), encrypt by share, store on multiple cloud service provider, decrypt share by recover, integrity check and original data item.

To detect file level and block level deduplication we are using hash calculation scheme with the help of various algorithms. To avoid collision attack we are using two algorithms at file level and two different algorithm at block level deduplication detection. This hash Calculation done by MD5, SHA and rabin fingerprint algorithm. File level Deduplication, Which discovers and removes redundant files to minimize storage demands or space. Block level Deduplication,

Which discovers redundancies between chunks or blocks of data in the file, and removes theses redundancy to reduce storage space. There are two types of block deduplication introduced, which are fixed size and variable size deduplication. In fixed size deduplication, fixed size blocks are checked for redundant copies while in variable sized deduplication variable sized blocks are checked for redundant blocks. After calculation hash values we are dividing single block of file in different secret shares with the help of secret share algorithm and distributing the secrets over different cloud storage sites to avoid insider attack, chosen distribution attack. To recover the secret we are using same secret share algorithm. In which by collecting various blocks in temporary file, and calculating hash again for proof of ownership scheme. We can recover the corrupt data or failure site, with the help of secret share algorithm. Also one of the more important feature of the proposed system is we can reconstruct the failure site, with the help of database and secret share algorithm without using traditional recovery ways like RAID etc.

*C. File level data deduplication*

Explanation: Fig. 1.3 is a File level deduplication. It comprises of file 1, file 2, same content, store only one file and give reference of file user file. There are various data deduplication methods or types. File level deduplication is one of them. If there are two files say file 1 and file 2 having same contents inside it. But these two files are used by two different users and uploads on client differently. If both the files are uploaded on the cloud then storage and bandwidth both are wasted so that file level deduplication is useful. Here only one copy of file is stored on the multiple cloud service providers and users get reference pointer to that particular file. It is the file level deduplication. Here two different hashing algorithms are used to check duplication. Hash value of complete file is calculated and compared with other file to check data deduplication. Two algorithms are used to avoid data collision attack.

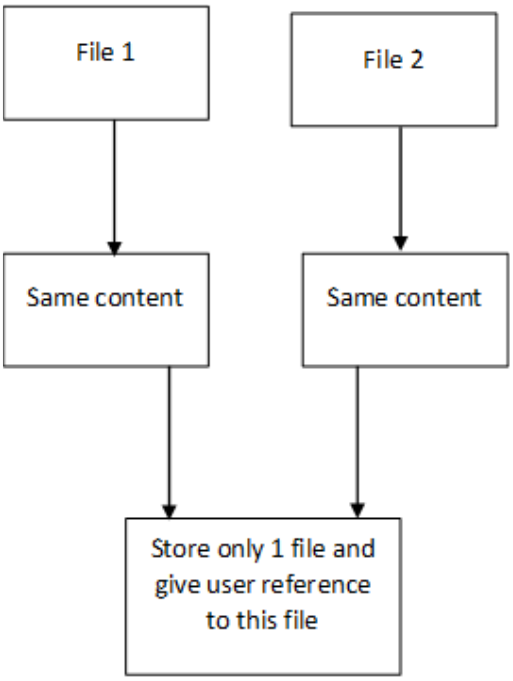


Fig.1.3 File level deduplication

*D. Block level data deduplication*

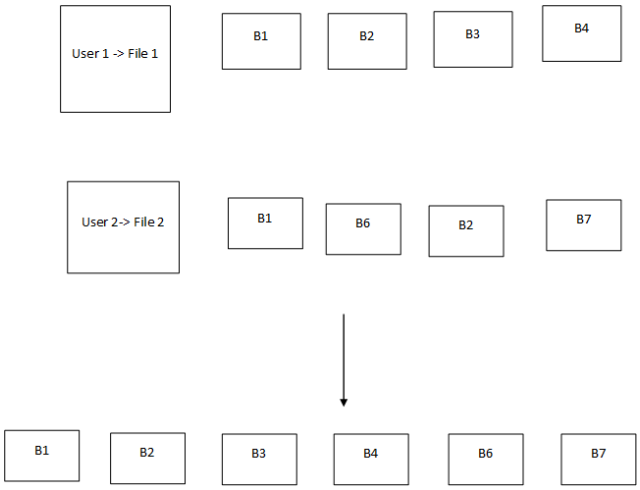


Fig.1.4 Block level data deduplication

Explanation: Fig. 1.4 is a block level deduplication. It comprises of user 1 file 1 and user2 file 2. In Block-level data deduplication technology data stream divided into blocks, check the data block, and determine whether it met the same data before the block (usually on the implementation of the hash algorithm for each data block to form a digital signature or unique identifier). If the block is unique and was written to disk, its identifier is also stored in the index; otherwise, the only deposit pointer to store the same data block's original location. This method pointer with a small-capacity alternative to the duplication of data blocks, rather than storing duplicate data blocks again, thus saving disk storage space. Hash algorithm used to judge duplicate data, may lead to conflict between the hash error. MD5, SHA-I, rabin fingerprint hash algorithm, etc. are checked against the data blocks to form a unique code.

Block level deduplication is divided into two parts:

1. Fixed Size data deduplication: In fixed size data deduplication file is divided into fixed size blocks and these blocks are checked with already stored blocks on the storage. If duplication found, then reference count is increased otherwise new data block is stored on the storage site.
2. Variable size data deduplication: In variable size data deduplication variable size blocks are created and checked for redundancy. Duplicate check is done by using hashing algorithms like MD5, SHA, rabin fingerprint etc.

*E. Secret share algorithm scheme working*

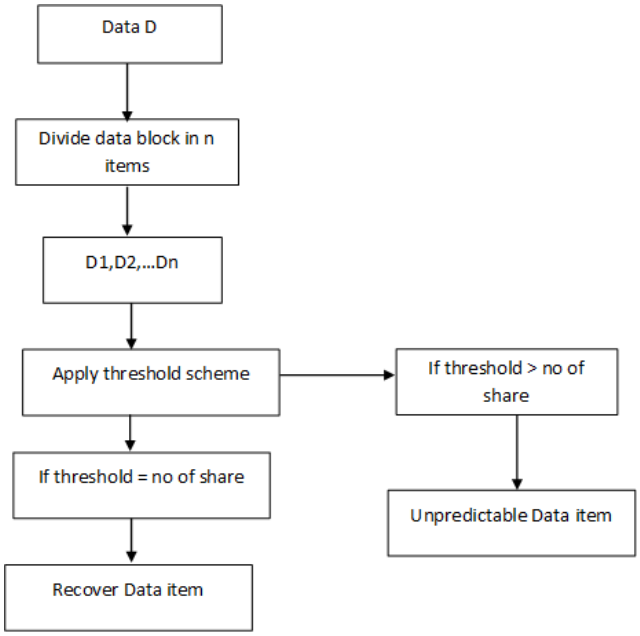


Fig. 1.5 Secret share algorithm scheme working

Explanation: Fig. 1.5 is a secret share scheme. It comprises of Data D, divide data block D in n items, Apply threshold scheme, and recover the data tem.

Secret Sharing Scheme is given by two algorithms: sharing (Share) algorithm and recovery (Recover) algorithm. Secret sharing technique, divides secret into multiple parts, each participant has given unique part of secret. To reconstruct the secret all or some parts of the secret must be present. To

reconstruct the secret all parts are necessary such condition is impractical, and therefore threshold scheme is introduced, which reconstructs the secret with the help of k(threshold) secret share. Share splits original message M into multiple pieces. As M is secret, Share is probabilistic algorithm because it introduces randomness. Recover is a deterministic algorithm which recreates the message M from some or all of the shares.

The Sharing Algorithm: Share (M) = (S1, S2, . . . ,Sn). The secrets S1, . . . , Sn are distributed securely among servers 1 through n.

The Recovery Algorithm: Recover (S1, S2, . . . , Sn ) = M.

If threshold of secret share algorithm equals to no of shares then original data item can be recovered. If less shares are there to reconstruct the secret that time the case is impossible. Secret cannot be predictable.

*F. Complete work flow diagram of Secure and data de-duplicated multi-cloud storage system (SDDMCSS)*

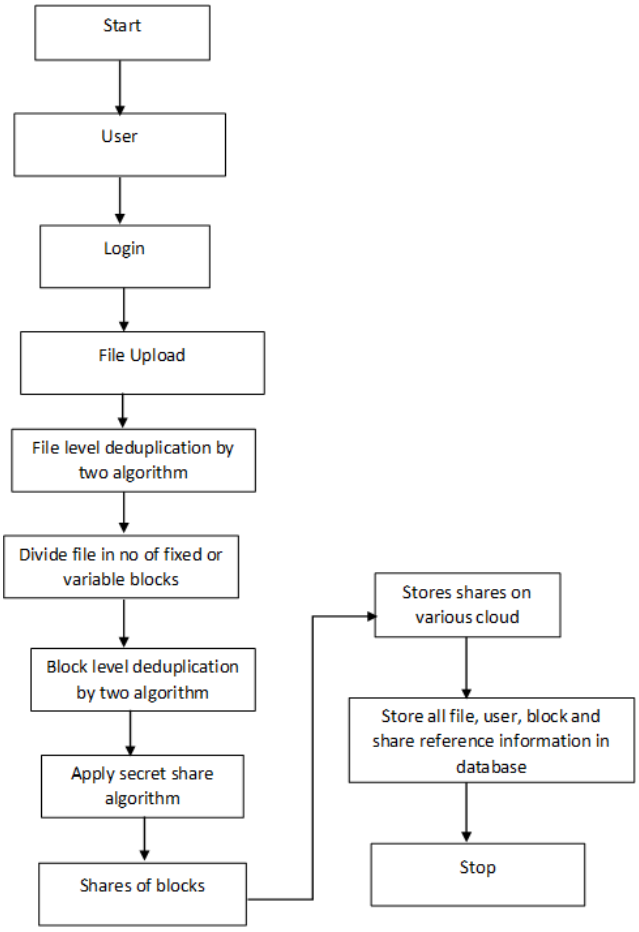


Fig. 1.6 Complete work flow diagram of Secure and data de-duplicated multi-cloud storage system (SDDMCSS)

Explanation: Fig. 1.6 Complete Work Flow Diagram of 'Secure and data de-duplicated multi-cloud storage system (SDDMCSS)'. Proposed system is reliable because we are using multiple cloud storage servers which having distributed structure. Here, distributed storage sites are used to save user data so even if any site fails user can get his original file without any fail. So obviously reliability of the system is more. The security requirement of reliability in deduplication means that the storage system can provide fault tolerance by using the means of redundancy. The system is required to detect and repair corrupted data and provide correct output for the users. Due to use of distributed storage site we recover data with the help of ant secret share algorithm like RSSS, shamir secret share etc. and database. Due to use of secret share algorithm we can recover the storage site easily without use of any backup methods or RAID array kind of types. In this mechanism due to use of distributed storage sites availability of data is more. Because according to secret share algorithm even if any site corrupted or failed still user file is available as per user demand. Data integrity is provided by two ways first by checking hash of file and second message authentication. Hash check is run by the cloud storage server during the file uploading phase, which is used to prevent the duplicate/cipher text replacement attack. If any adversary uploads a maliciously-generated cipher text such that its tag is the same with another honestly-generated cipher text, the cloud storage server can detect this dishonest behaviour. Thus, the users do not need to worry about that their data are replaced and unable to be decrypted. Message authentication check is run by the users, which is used to detect if the downloaded and decrypted data are complete and uncorrupted or not. This security requirement is introduced to prevent the insider attack from the cloud storage service providers. As shown in Fig. 1.6 'Secure and data de-duplicated multi-cloud storage system (SDDMCSS)', workflow takes place step by step. Initially user uploads any file over multiple cloud. File level deduplication check has been done by two different algorithms. These two algorithms are hashing algorithms which calculates hash value of file and checks this value with already stored file, if match found then file will not be stored actually on the storage servers instead of that reference to that particular file is given to the user internally. After file level check file is checked for block level deduplication, here file is divided in to fixed size or variable size blocks and checked for duplicate data. Here also two hashing algorithms are used to check duplicate data and avoid data collision attack. After block level deduplication actually block storage operation gets start. Block is stored on the multiple cloud service provider through secret share algorithm. It divides the blocks in no of shares and gives some threshold to recover. It divides the block in such a way that single part of that block is unpredictable. As it uses threshold scheme to recover original data. If no of shares distributed over multiple cloud are less than defined threshold then original data cannot be discovered, if threshold equal to no of shares then original data is predictable. After data recovery again hash value calculated and checked for data integrity purpose. In this way blocks of

files are recovered then complete file is regenerated from multiple blocks. Through present invention security, integrity and availability of user file has been done secure way.

*G. Complete workflow diagram of data reconstruction of failure storage site*

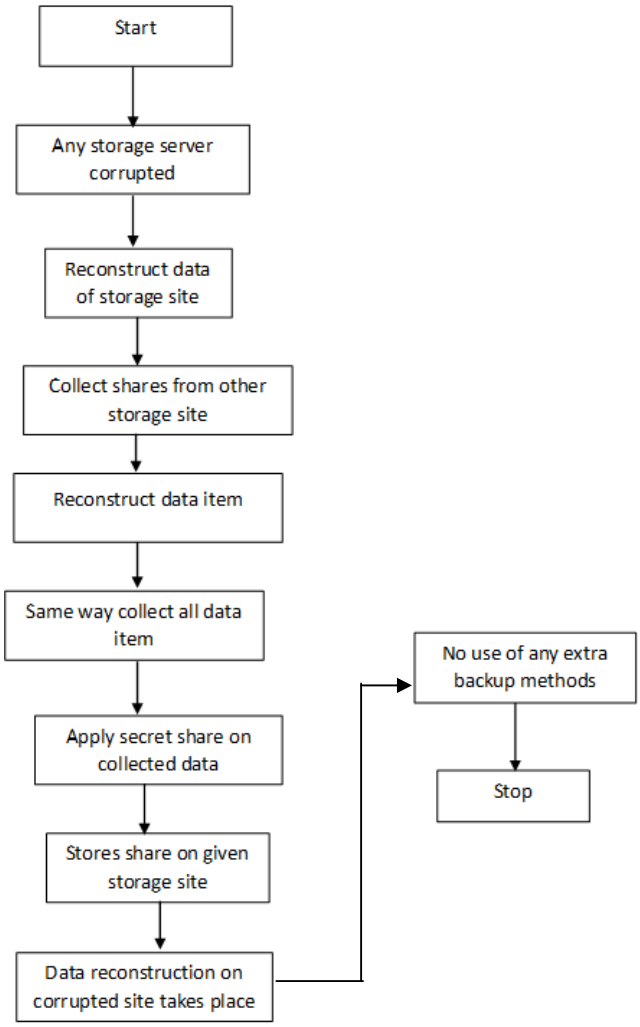


Fig. 1.7 Complete work flow diagram of reconstruction of failure storage site

Explanation: Fig.1.7 is a complete workflow diagram of Data reconstruction on corrupted storage site. Consider a situation what happen if complete cloud storage site on which user data stored is corrupted or hacked? The solution to above problem is there are multiple backup methods are available to recover data. Instead of using traditional backup methods one can reconstruct a data using the secret share technique. If complete site on which shares are stored is corrupted due to some reason then start the data reconstruction operation back end. Initially shares from other sites are collected and checked for data integrity by calculating hash value of those shares. After that a complete data block is generated from other shares except corrupt share using threshold scheme of secret share algorithm. Original data construction has been done then, again apply secret share to the correct block and divide share on multiple cloud storage server. Here deletion of previous shares has been done simultaneously. New shares are stored on these storages having same hash value and same block. This is the property of any secret share algorithm. By using this property we can generate corrupted storage site easily without use of backup methods.

In this way 'Secure and data de-duplicated multi-cloud storage system (SDDMCSS)' performs various levels of data deduplication in a secure way to minimize bandwidth and storage space of multiple cloud service providers.

## V. Novel feature & Analysis of Secure and data de-duplicated multi-cloud storage system (SDDMCSS)

Data set contains set of data record which is used to test the functionality of the system. To find saved storage space and network bandwidth following data set has been extracted. Here different files are judged for storage space and network bandwidth with duplicate, half duplicate and unique type. We can use any kind of file to upload .Following table describes how proposed system can save storage and network bandwidth in duplicate and half duplicate files. In table 1.1 storage space and network bandwidth saving is represented in the percentage form according to status of file.

Table 1.1 File saving in SDDDMCSS Table

| File | Duplicate check | Required Storage space | Required bandwidth | Saved space and bandwidth |
|---|---|---|---|---|
| Text file | Duplicate | 0% | 0% | 100% |
| Audio | Half duplicate | 50% | 50% | 50% |
| video | duplicate | 0% | 0% | 100% |
| Image file | Unique | 100% | 100% | 0% |

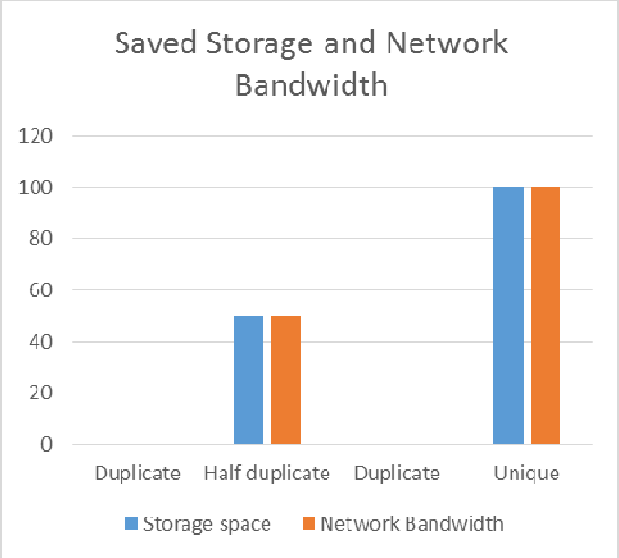


Fig. 1.8 Saved storage space and network bandwidth

Fig 1.8 has been drawn by considering values in Table 1.1 As shown that with the help of proposed system one can save maximum amount of storage space and network bandwidth. Here proposed system considered duplicate and half duplicate concept to store a file. As shown in fig 1.8, duplicate files are not stored actually on the multiple cloud because those files are already stored on the cloud. Those files are duplicate files. Half duplicate means some part of the file is already present, so unique part only stored on the cloud. Such a system is useful not only for cloud service provider but also normal user. Because it saves storage space of cloud service provider and bandwidth of user.

*A. Advantage(s) of Secure and data de-duplicated multi-cloud storage system (SDDMCSS):*

1. ''Secure and data de-duplicated multi-cloud storage system (SDDMCSS)' does data deduplication to store single copy of same data files.
2. 'Secure and data de-duplicated multi-cloud storage system (SDDMCSS)' is a software system stores single copy of file along with maximum security and reliability.
3. 'Secure and data de-duplicated multi-cloud storage system (SDDMCSS)' provides security, confidentiality, reliability and availability to user data.
4. 'Secure and data de-duplicated multi-cloud storage system (SDDMCSS)' uses multiple cloud storage servers to store data blocks of files.
5. 'Secure and data de-duplicated multi-cloud storage system (SDDMCSS)' storage space and data bandwidth is stored.
6. 'Secure and data de-duplicated multi-cloud storage system (SDDMCSS)', performs at smallest level of data deduplication to save storage space.
7. 'Secure and data de-duplicated multi-cloud storage system (SDDMCSS)', single part of data block is unpredictable to any user.
8. 'Secure and data de-duplicated multi-cloud storage system (SDDMCSS)', uses threshold scheme to store and retrieve original data blocks.
9. 'Secure and data de-duplicated multi-cloud storage system (SDDMCSS)' used anywhere to store unique copy of data or files.
10. 'Secure and data de-duplicated multi-cloud storage system (SDDMCSS)', data blocks stored on multiple cloud servers, so single server data blocks cannot identify original data.
11. Even if whole storage server corrupted, we can easily generate the data files on the corrupted server without any maximum storage facility as like RAID array method.
12. Proposed system can be run on any system.

**Conclusion**

'Secure and data de-duplicated multi-cloud storage system (SDDMCSS)', secure user data without using traditional encryption mechanism. The construction is proposed to work on the file-level and fine-grained block level data deduplication to minimize the storage space and bandwidth. The proposed deduplication system uses the secret sharing scheme for storing the user data. In secret sharing, Secret is divided into multiple parts or shares, to reconstruct the secret some parts of secret must be available. Single share of secret is difficult to predict the original secret. Proposed system achieves confidentiality, Integrity and reliability by removing various attacks and provides secure distributed deduplication system in the cloud environment. This mechanism achieves strong security against insider and outsider attack, minimum storage space as well as saves network bandwidth. This mechanism itself is a cloud infrastructure with great performance as compared to the traditional deduplication and can be applied to the current cloud service provider to achieve the various feature implemented in the proposed mechanism. The proposed mechanism gives minimum overhead as compare to the traditional deduplication system. Proposed system recovers failure storage site with the help of database and Shamir secret sharing scheme instead of traditional recovery or backup methods (RAID or any other). Here the smallest shares of data items are stored on multiple cloud storage providers rather than single cloud service provider to provide more security to user data and achieve availability of data.

## References


[1] Ashutosh Saxena, Nitin Singh Chauhan, Sravan Kumar Rondla, "System and method for verifying integrity of cloud data using unconnected trusted device," Patent No US 9641617 B2, published date May 2, 2017.

[2] Bogdan Nicolae, " Cost-effective IAAS (infrastructure-as-a-service) cloud storage based on adaptive virtual disks (AVD)," Patent No US 9442669 B2, published date 13 Sep 2016.

[3] Rahul V. Auradkar, Roy Peter D'Souza, " Secure and private backup storage and processing for trusted computing and data services," Patent No US 20100318812 A1, published date 16. Dec. 2010

[4] Jonathan M. Barney, Cataldo Mega, Edmond Plattier, Daniel Suski, "Method and system for managing security in a computing environment," Patent No US 9560019 B2, published date 31 Jan 2017.

[5] Jens-Matthias Bohli, Ghassan KARAME ," A method for storing of data within a cloud storage and a cloud storage system" Patent no WO 2016026537 A1, published date Feb 25, 2016.

[6] David Lanc, Lu Fan, Lachlan MACKINNON, Bill BUCHANAN ," Resilient secret sharing cloud based architecture for data vault" Patent no US 20170005797 A1, published date 5 Jan 2017.

[7] Ashutosh Saxena, Nitin Singh Chauhan, Sravan Kumar Rondla ," System and method for verifying integrity of cloud data using unconnected trusted device" Patent no US 9641617 B2, published date May 2, 2017.

[8] Prof.V.N.Dhawas, Pranali Juikar ,Neha Patekar ,Neha Lendghar,Sushant Vartak , "A Secured Cost Effective Multi -Cloud Storage in Cloud Computing" in International Journal of Scientific & Engineering Research, Volume 4, Issue 5, May -2013

[9] Jin Li, Xiaofeng Chen, Xinyi Huang, Shaohua Tang and Yang Xiang Senior Member, IEEE and Mohammad Mehedi Hassan Member, IEEE and Abdulhameed Alelaiwi Member, IEEE , "Secure Distributed Deduplication Systems with Improved Reliability." in IEEE Transactions on Computers Volume: PP Year: 2015.

[10] J. Li, X. Chen, M. Li, J. Li, P. Lee, andW. Lou, "Secure deduplication with efficient and reliable convergent key management" in IEEE Transactions on Parallel and Distributed Systems, 2014, pp. vol.25(6), pp. 16151625.

[11] M. Bellare, S. Keelveedhi, and T. Ristenpart, "Dupless: Serveraided encryption for deduplicated storage." in USENIX Security Symposium, 2013.

[12] Mihir Bellare, Sriram Keelveedhi and Ristenpart, "Message-locked encryption and secure deduplication." in EUROCRYPT, 2013, pp. 296312.

[13] J. Stanek, A. Sorniotti, E. Androulaki, and L. Kencl, "A secure data deduplication scheme for cloud storage" in Technical Report, 2013.

[14] Shamir, Adi, "How to share a secret." in Communications of the ACM 22 (11):612613

[15] Prerna Lahane, Prof. Sarika Bodake, "Deduplication of Distributed Cloud Storage by Improving onfidentiality, Integrity and Reliability" in International Journal of Innovative Research in Computer and Communication Engineering, Vol. 4, Issue 6, June 2016



Prof. Dr. Biplab Kumar Sarkar, Post.Doc. (Singapore), PhD.(CS)(IIT-BHU), M-Tech(CS) IIT-BHU, B-Tech(CS). Member. IETE, ISTE, SMU, IIHT, Global R/D, GEH. He is having 15 years of academic as well as industrial experience. He holds various positions like He has worked on many funded project. Since 2003 total approx.1200 Cr. Governmental/ Non-Governmental project completed. He is having 03 patent filled and 03 is under provisional patent process. He is associated with 05 various universities. He has published 03 books in Engineering field and 02 is under process .He has published 20 papers in national/International journals and conference.

Prof. Dr. Bej Raj Singh Patel, Post.Doc. (Japan), PhD.(CS), M-Tech(CS), B-Tech(CS). Member. IETE, ISTE, SMU, IIHT, Global R/D, GEH. He is having 12 years of academic as well as industrial experience. He holds various positions like He has worked on many funded project. Since 2005 total approx.900 Cr. Governmental/ Non-Governmental project completed. He is having 03 patent filled and 03 is under provisional patent process. He is associated with 05 various universities. He has published 03 books in Engineering Field and 02 is under process .He has published 15 papers in national/International journals and conference.