Secure Data Deduplication Scheme

Prof. Ajay Dhruv[1] Snehal Shete[2] Sonal Kajrolkar[3] Subodh Ambavane[4]
[1,2,3,4]Vidyalankar Institute of Technology, Mumbai, India.

Abstract— In cloud storage utility, the de-duplication technique is frequently used to degrade the storage and bandwidth by eliminating duplicate data and storing just one copy of it. Convergent encryption has been affirmed to perform secure encrypted de-duplication in the mix user scenario. De-duplication is extremely helpful when there are many numbers of users stores the similar kind of data to the cloud storage, but the problem arises when it comes to security. Thus, security has been amended in the purpose system. To make data management scalable in cloud computing, De-duplication has been a well-known technique for eliminating duplicate copies of repeating data in storage. The technique is used to improve storage utilization and can also be applied to network data transfers to reduce the number of bytes that must be sent. As opposed to keeping various data copies with a comparative substance, De-duplication takes out overabundance data by keeping only a solitary physical copy and implying other dull data to that copy. De-duplication can take place at either the file level or the block level. For file level De-duplication, it eliminates duplicate copies of the same file. De-duplication can also take place at the block level, which eliminates duplicate blocks of data that occur in non-identical files and has attracted more and more attention recently. Data De-duplication is a specialized data compression.

Key words: Data De-duplication, Cloud Computing, Cloud Storage, Data Security

I. Introduction

Data De-duplication is efficient technique to handle these large duplicate data. With the quick gain in cloud computing large number of data are being centralized into the cloud in a successively increasing manner. According to the previous analysis 70% of data in the world is a copy, and there is no doubt that the size of data will be enormously extend hence one of the hazardous challenge will be the efficiently managing of this data. To reduce resource expenditure, many cloud storage services adopt de-duplication technique, such as Drop box, Google drive, Microsoft one drive, where cloud server stores just one copy of duplicate data and provide links that referred to the original copy instead of storing all redundant data irrespective of number of clients ask to store data.

Consequently, effective organization of huge scale of data imposes a critical challenge and sharing the data with any user arises the issue of security, as clients are more concern about there private data. A lot of ideas and research has been done in this field. In addition numerous approaches has been used for encryption and organization of data .this paper represents the survey on various methodologies like randomized data encryption(RCE), interactive randomized data encryption(IRCE), convergent encryption(CE) used for de-duplication scheme..

II. Literature Survey

This section presents the survey on secure data de-duplication in cloud storage.

A. Duples Server-Aided Encryption for Duplicated Storage Dupless:

Server aided encryption for duplicated storage for distributed storage specialist co-op like Mozy, Drop box, and others perform de-duplication to Message lock encryption is used to resolve the problem of clients encrypt their file.Customers scramble under message-based keys got from a key-server by means of an unaware PRF convention in dupless server. It permit customers to store scrambled information with a current administration, have the administration happens de-duplication on their on the part, but accomplishes solid classification ensures. It show that encryption for duplicated storage can successfully reach wanted execution and space investment funds near that of utilizing the capacity benefit with plaintext data.

Characteristics:

1) More Security.
2) Easily-deployed solution for encryption that supports De-duplication.
3) User Friendly: Use command-line client that support both Drop box and Google Drive.
4) Resolve the problem of message lock Encryption.

B. Evidences of Ownership in Remote Storage Systems:-

It stores only the single copy of the duplicate data. Customer side de-duplication tries to recognize de-duplication chance as of now at the customer and spare the data transmission of transferring duplicates of existing documents to the server. To overcome the attacks Shai Halevi1, Danny Harnik, Benny Pinkas, and Alexandra Shulman-Peleg proposesthe Proof of ownership which lets a customer effectively demonstrate to a server that that the customer keep a record, rather than just some short information about it present solutions based on Merkle trees and specific encodings, and analyze their security.

Characteristics:

1) To distinguish the assaults that endeavor customer side de-de-duplication. Proofs of possession give the thorough security.
2) Rigorous efficiency requirements of Peta-byte scale storage systems.

C. A Secure De-Duplication with Efficient And Reliable Convergent Key Management: -

Data de-duplication is a used for ousting duplicate copies of data, and has been for the most part associated in appropriated stockpiling to lessening storage space and in addition exchange transmission limit. Promising as it seems to be, a showing up test is to fulfill secure de-duplication in distributed storage.

Techniques:

1) Key management
2) Convergent Encryption

D. *Twin Clouds:*

An Architecture for Secure Cloud Computing proposed architecture for secure outsourcing of data and arbitrary computations to an untrusted commodity cloud. In come towards, the user communicates with a trusted cloud Which scrambles and additionally confirms the information put away and operations happened in the untrusted cloud .It divide the computations such that the trusted cloud is used for security-critical operations in the less time-fundamental setup organize, while request to the outsourced data are processed in parallel by the fast cloud on encrypted data.

E. *Private Data De-duplication Protocols in Cloud Storage:*

In this paper, there are two classifications of information de-duplication system, and expand the blame tolerant computerized signature conspire proposed by Zhang on looking at repetition of pieces to accomplish the information de-duplication. The proposed plot in this paper decreases the distributed storage limit, as well as enhances the speed of information de-duplication. Furthermore, the signature is computed for every uploaded file for verifying the integrity of files.

III. METHODOLOGY

The linear working of our project can be explained in the following manner

STEP 1:

User

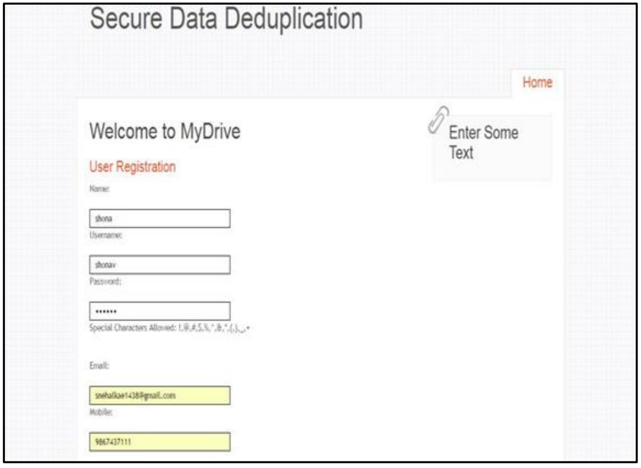
User will login into the system

In this system the user will first login into the system to use the secure data de-duplication scheme. The user will fist upload and download their files on the cloud storage.

STEP2:

Registration:

If the user is using our system for the first time then he has to register him on our system to get the benefit of our cloud storage system. If user has already registered then he can login into the system directly.


Secure Data Deduplication

Home

Welcome to MyDrive

Enter Some Text

User Registration

Name: shona

Username: shonav

Password: ••••••

Special Characters Allowed: !,@,#,$,%,^,&,*,(,),_,+

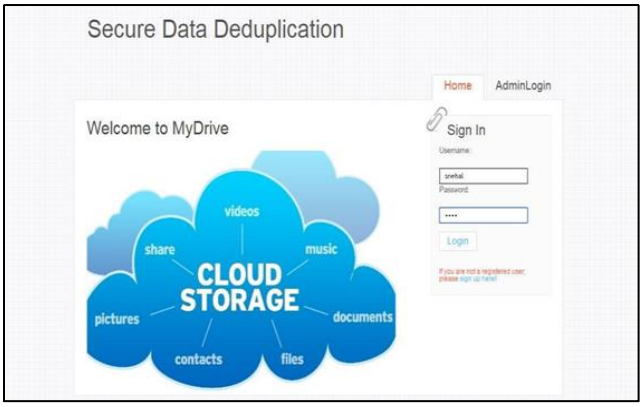
Email: snehalkae1438@gmail.com

Mobile: 9867437111

Fig. 1: Registration

STEP 3:

Login


Secure Data Deduplication

Home AdminLogin

Welcome to MyDrive

Sign In

Username: snehal

Password: ••••

Login

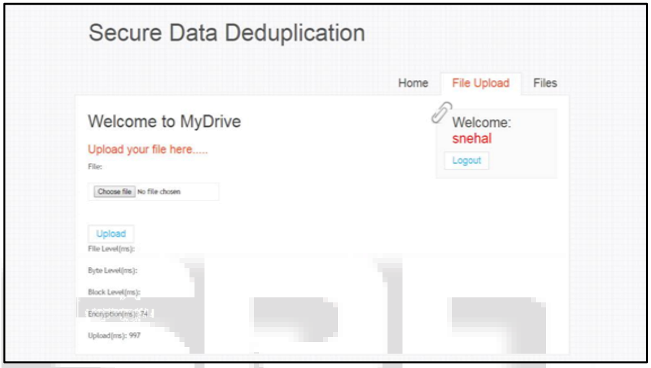
If you are not a registered user, please sign up here!

Fig. 2: Login

After successful registration user can log into the system and can upload and download their files on the system. The user can access our system completely after the login.

Step 4:

Upload


Secure Data Deduplication

Home File Upload Files

Welcome to MyDrive

Welcome: snehal

Logout

Upload your file here.....

File:

Choose file No file chosen

Upload

File Level(ms):

Byte Level(ms):

Block Level(ms):

Encryption(ms): 74

Upload(ms): 997

Fig. 3: Upload

The user can upload its file on the system if files content is not duplicated with the already existing files on the system then the user can upload its new file on the system.

If the files content is matched with existing file then the system will not allowed to upload these file. If you want to upload these file then you can make changes to the file and upload it on the system.

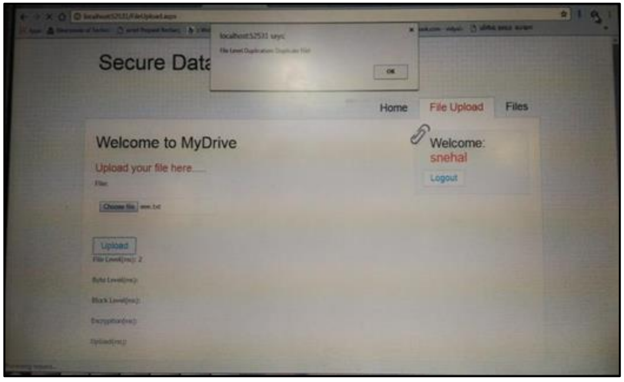
Step 5:

Block Level Deduplication:

In Block level each file is divided in the blocks and then only unique block will be stored. The file is divided into 5 blocks and the content in each block is checked and hash value is generated according to that. The encryption is performed on the file by using AES rijendial algorithm.

Step 6:

Check Duplication:


localhost:52531 says:

File Level Duplication: Duplicate File

OK

Secure Data

Home File Upload Files

Welcome to MyDrive

Welcome: snehal

Logout

Upload your file here.....

File:

Choose file www.txt

Upload

File Level(ms): 2

Byte Level(ms):

Block Level(ms):

Encryption(ms):

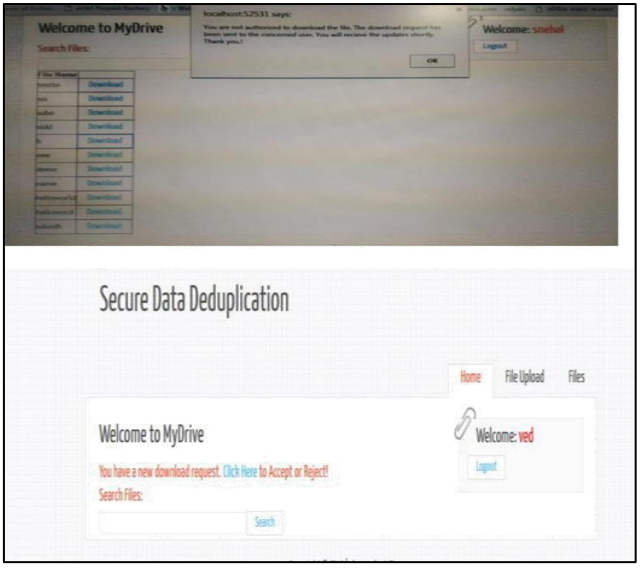
Upload(ms):

Fig. 4: Check Duplication

If the content of the file is 80% and more than that is same with existing file then the file is not uploaded. If the data of the file is not duplicated then the file is uploaded successfully on the system.

Step 7:

Access Request to Owner



Welcome to MyDrive

Search Files:

File Name	
tendto	Download
ss	Download
subo	Download
sidd	Download
b	Download
one	Download
demo	Download
name	Download
helloworld	Download
helloword	Download
subodh	Download

localhost:52531 says:

You are not authorized to download the file. The download request has been sent to the concerned user. You will recieve the updates shortly. Thank you.!

OK

Welcome: snehal

Logout

Secure Data Deduplication

Home File Upload Files

Welcome to MyDrive

You have a new download request. Click Here to Accept or Reject!

Search Files:

Search

Welcome: ved

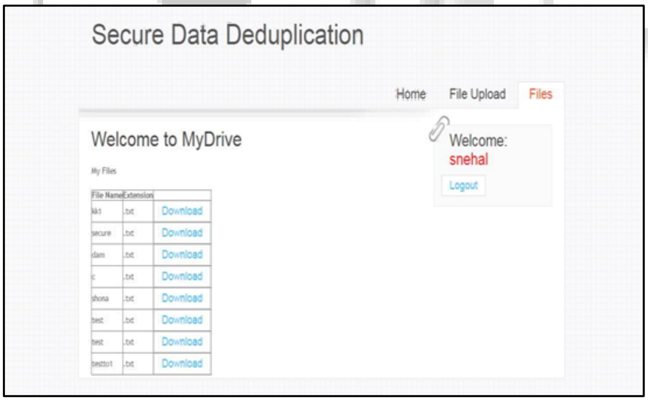
Logout

Fig. 5: Access Request to Owner

The owner is the one who is uploading the new file for the first time on the system. The another user whose file is duplicated can send the request to the owner for accessing the file then the owner can either accept the request or deny it. If the owner has accepted the request of the user then the user can access the file on the system.

Step 8:

Download



Secure Data Deduplication

Home File Upload Files

Welcome to MyDrive

My Files

File Name	Extension	
kk1	.txt	Download
secure	.txt	Download
dam	.txt	Download
c	.txt	Download
shona	.txt	Download
test	.txt	Download
test	.txt	Download
testto1	.txt	Download

Welcome: snehal

Logout

Fig. 6: Download

After access given by the owner the user can download the file from the system and can use it.

IV. System Flow

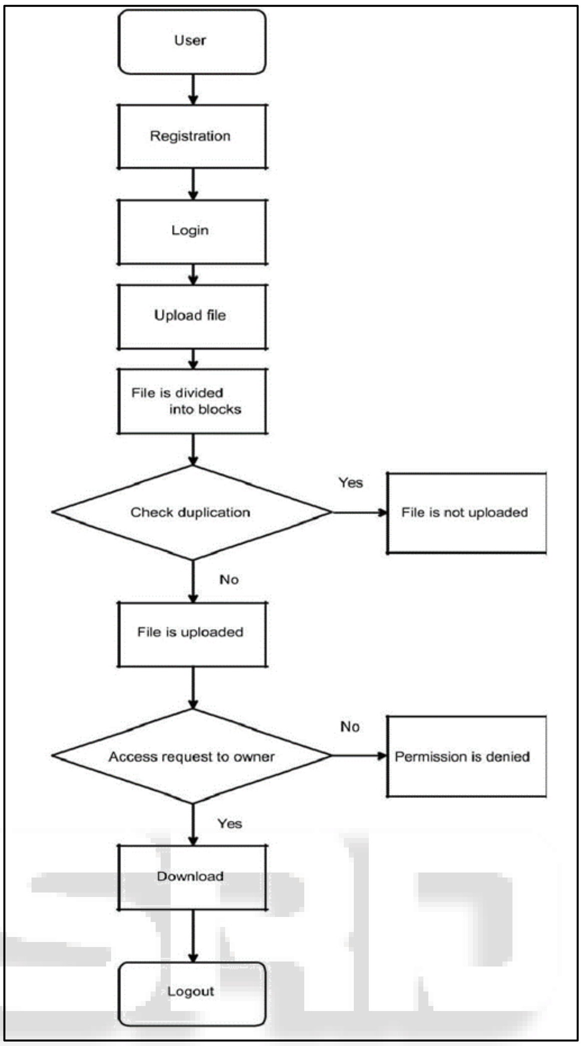


Fig. 4.1:

V. Algorithms Used in Preprocessing

The AES Rijndael algorithm is used for the secure data de-duplication scheme. The steps of Rijndael algorithm are following:

1) Byte Substitution: Each 8-bit byte in the state is reversibly mapped into another byte. An S-box is a lookup table that maps n bits to m bits. It's the only part of the cipher that is non-linear, and considered the most important part of the algorithm.
2) Shift Row: Here, rows are shifted over four different offsets. Row 0 is unmoved; Row 1, Row 2 and Row 3 are rotated left 1, 2 and 3 bytes respectively.
3) Mix Column: In this step, the bytes in columns are linearly combined and Matrix multiplication is performed.
4) Add Round Key: This is the final step where the sub key is XORd in for the current round.

VI. Conclusion

In this paper, the thought of approved information de-duplication was proposed to ensure the information security by including differential privileges of users in the duplicate check. We also presented a few new de-duplication developments supporting approved copy check in crossover cloud architecture, in which the duplicate-check tokens of files are generated by the private cloud server with private keys. Security examination exhibits that our plans are secure in wording of insider and outsider attacks specified in the proposed security model. We showed that our approved copy check plot acquires negligible overhead contrasted with focalized encryption and network transfer. Moreover, the encryption of the file to store in the cloud is done attribute based. The cloud does not know the identity of the user who stores information, but only verifies the user's credentials.

Enter appropriation is done decentralised and furthermore conceal the traits and get to strategy of a client. For more security, session grid for password authentication is used. This helps to prevent the shoulder surfing attack.

REFERENCES

[1] Enhanced Secure Thresholded Data Deduplication Sc heme for Cloud Storage"Jan Stanek; Lukas Kencl"; Year: 2016
[2] An Improved velocity Energy-efficient and Link-aware Cluster-Tree baseddata collection scheme for mobile networks "S. Gopalakrishnan; P. Mohan Kumar" Year: 2016
[3] A Realistic Lightweight Anonymous Authentication Protocol For Securing Real-time Application Data access in Wireless Sensor Network "Prosanta Gope; Tzonelih Hwang" Year: 2016
[4] Secure and Efficient Data Transmission for Wireless Mesh Networks "Luis Cabrera; Tuan Tran; Scott Cordle; Emmanuel Udoh" Year: 2015