An Efficient Framework and Technique to Maintain Privacy and Ovrcome Duplication Overhead In Cloud

[1]Zarka khan, 2 Mr. Anwar Ahmed Sheikh,
[1]M. Tech Scholar, Department of Computer Science, Integral University, Lucknow, India.
[2]Department of Computer Science, Integral University, Lucknow, India.
Corresponding Author: Zarka Khan

ABSTRACT*- Social networking sites a few years back was just used as a medium for communication for making new friends etc. If we compare today's current situation from years back, we'll notice a huge difference in terms of how these sites maintain theirdata. Nowadays social networking sites stores their data online Cloud storage can provide the benefits of greater accessibility, maintaining privacy and reliability; rapid deployment; strong protection for data backup, archival and disaster recovery purposes.Today our modern era is significantly dependent on Internet and online storage. Our proposed Technique is to maintain memory efficiently by not storing the same images again in cloud. In this paper, a novel approach is proposed. A image once uploaded is allotted an id which is unique and the image occupies the space in cloud storage and if any similar image is uploaded again then the reference of the similar stored image is used to upload that image and it does not occupy any memory in cloud. The proposed approach in this paper improves memory efficiency and overcomes deduplication overhead and maintains the privacy of the user.*
Keyword- *Cloud Storage, Web Hosting, Cloud Computing, Data Redundancy, Data Deduplication, RSA Algorithm, MPODO.*

I. INTRODUCTION

Data is increasing day by day on social networking sites it has become important to provide users with easy environment so that users can share their contents more efficiently and memory can be used efficiently avoiding data redundancy. Cloud technology is changing the way of seeing storage needs and computational powers. Cloud technology in social sites is providing more flexibility to share data as well as it uses deduplication technique to overcome data redundancy and maintain memory efficiently. Once data deduplication is done it can reduce the communication and storage overheads in the cloud storage services. There are many data deduplication schemes, but they can either resist brute-force attacks or ensure the efficiency and data availability but cannot satisfy both the conditions. There is no such existing scheme which achieves accountability, in the sense of reducing duplicate information discloser. Henceforth, three-tier-cross-domain architecture is used that can maintain privacy and overcome duplication overhead in cloud storage (hereafter referred to as MPODO). MPODO achieves deduplication along with both privacy-preserving and data availability, and resists brute-force attacks. Although, data deduplication provides a lot of benefits, there are some security and privacy concerns as the sensitive data is susceptible to many kinds of attacks. So, to overcome these issues a referential key generation technique called RSA is also included in this approach.The most important thing is that is taken into consideration in this paper is accountability to offer better privacy assurance in comparison to other existing schemes. MPODO can outperforms existing competing schemes, in terms of computation, communication and storage overheads. The time complexity of duplicate search over cloud in MPODO is logarithmic. In this the comparison of the image is done in a zone.

1.1 Cloud Computing

In cloud user is not aware about where their data is going to be stored. Cloud is basically defined as a term which can be stored anywhere in the world. Cloud Computing refers to accessing, configuring and manipulating the software and hardware resources remotely. It provides user with online infrastructure, data storage and application. Cloud supports pay per use facility and platform independency, as software need not be installed locally on the user's PC. Data stored should be encrypted form to provide security. Cloud computing is as

important as water and air which is changing trend very rapidly day by day. It is providing more flexibility to share data on social sites and is changing the way of seeing storage need.

1.2 Cloud Storage
Cloud storage can be defined as the storage of data online in cloud. It provides facilities like large reliability and accessibility; rapid deployment; security for data backup, archival and disaster recovery purposes; and the overall cost is decreased but of the pay per use facility and can easily manage and maintain expensive hardware. There are large numbers of cloud storage providers on the Web, and more are popping up all the time. Apart from providing storage, the amount of storage each company offers to its clients is also growing rapidly.

1.3 Data Deduplication
It is a data compression technique that eliminates duplicate copies of similar repeating data. This technique is mostly used in cloud server to reduce space in server. Data is encrypted using different encryption technique before storing it on server. Along with saving the data storage it also reduces the amount of bandwidth of data transfer. Data deduplication works by eliminating the repeated data and it stores only the first unique instance of any data. If the user tries to store the similar data again then only the reference id is created to originally store the data rather than storing the replica.

1.4 Cloud Hosting:
Cloud hosting refers primarily to the use of virtual hardware, network, storage and composite solutions from a cloud vendor. It is enabled through virtualization, whereby the entire computing capacity of an infrastructure or data centre is distributed and delivered to multiple users simultaneously. The user uses underlying infrastructure to host its own applications, services and data. For example, a physical server can be virtualized and consolidated to host several cloud servers, all sharing the processor, memory, storage, network and other resources. Moreover, cloud hosting may also combine the capacity of several servers to provide a single cloud hosted server. Some of the cloud hosted solutions are cloud servers, cloud desktops, cloud storage and more.

1.5 Web Hosting Service:
A web hosting service is a type of Internet hosting service that allows individuals and organizations to make their website accessible via the World Wide Web. Web hosts are companies that provide space on a server owned or leased for use by clients, as well as providing Internet connectivity, typically in a data center. Web hosts can also provide data center space and connectivity to the Internet for other servers located in their data center.

II.   LITERATURE REVIEW:
Rongzhi Wang (2017) [1] in his paper "Research on data security technology based on cloud storage" proposed a secure storage scheme based on Tornado codes (DSBT) by combining the technique of symmetric encryption and erasure codes. Based on Cassandra this paper implements a secure cloud storage prototype. It focuses mainly on how timely users are informed that their data is intact; how to restore data if it is not in good condition; how efficiently the key can be implemented to solve the problems in the cloud storage environment.
Yan, Wang, Li and Vasilakos [2] in paper "Encrypted Data Management with Deduplication in Cloud Computing" proposed a scheme for deduplication which provides secure access using ABE (attribute-based encryption). SHA1 is used for hash function by the system, for deduplication it uses CP_ABE, for symmetric encryption it uses AES and RSA for public key Cryptography. It supports data deletion, updation and deduplication with low operational and implementation cost as third party is not involved for key generation. Drawback of this scheme is that it takes more time for key generation.
Alzain, Ben and Eric [3] in paper "A New Approach Using Redundancy Technique to Improve Security in Cloud Computing" proposed a new model called (MCDB) which uses multi-clouds instead of single cloud service provider. It also uses the Shamir's secret sharing techniques and it also uses TMR (triple modular redundancy) with sequential method to improve the system reliability and enhance privacy and security.
Shobana, Shantha, Sridevi and Leelavathy [4] in their paper "De-Duplication of Data in Cloud" the data is encrypted using AES (Advance Encryption Standard) before being stored. System uses a effective user authentication using fingerprint feature extraction, image based authentication during file upload/download, eliminating repetition of data in cloud server and is implemented through multiple cloud storage. The convergent encryption technique is used to encrypt the data before outsourcing. It is extensively used in cloud storage to save bandwidth and minimize the storage space. The drawback is that it uses multiple cloud storage and splits the files and database of a specific user before storing.

III.   PROPOSED APPROACH

Step 1:

The existing user can login to his site with the help of his username and password. A new user can also register with login details like name, email and password etc. Each user is provided with a unique ID and password which is only known to the authorized user so that any unauthorized user can't access the account. Once the user is successfully logged in, he can upload the image on his account.

Step 2:

Cloud dashboard is used is used for storage which can be build and accessed on a web browser. It has key advantages over on-premise software as it is easy to deploy, requires little to no IT support and can be accessible to multiple device.

Step 3:

In this module we are going to check whether the image is already present in the cloud database or not. The image is first converted into JPEG format for comparison. Once the image is uploaded before posting and storing, it is compared with the images present in the database to avoid duplication. The image RGB format is converted into video-based encoding format which separates the grayscale (luminance)and colour (chrominance). Then both luminance and chrominance information are transformed from spatial into frequency domain. Both the information is divided into square (8*8) block and then 2-D Discrete Cosine Transform (DCT) is applied to each block. The transform produces an array of coefficient that represents each block of data in the frequency domain. Each block is then subjected to quantization cosine transform coefficient is divided by the corresponding element in the default quantization matrix and the value is rounded. The result compromises of large number of zeros. A zigzag pattern is employed that progresses from low frequency to high frequency. Because high-frequency terms are eliminated therefore any block compromises of large run of zeros at the end. Thus, the amount of space required to represent the block is reduced. Once compression is done of both the images then they are then bit by bit comparison using string matching technique.

Step 4:

If the image is unique then it occupies space in the cloud and is uploaded. If the image already exists on the cloud server then using the RSA technique the referential key is created of the preloaded image and the key is further encrypted using RSA encryption to maintain privacy. Now instead of storing the image the reference of the preloaded image is stored in the record of upload column and the reference is further used to upload the image rather than uploading the actual image.

The flow chart and system architecture are shown in the figure below:
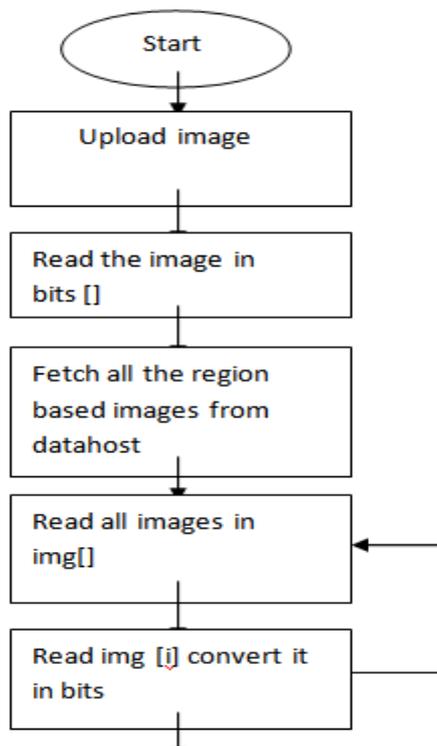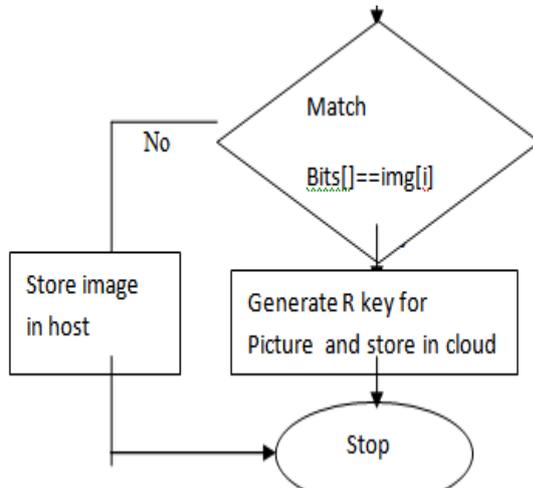
**Figure 1: Flow diagram of proposed approach**
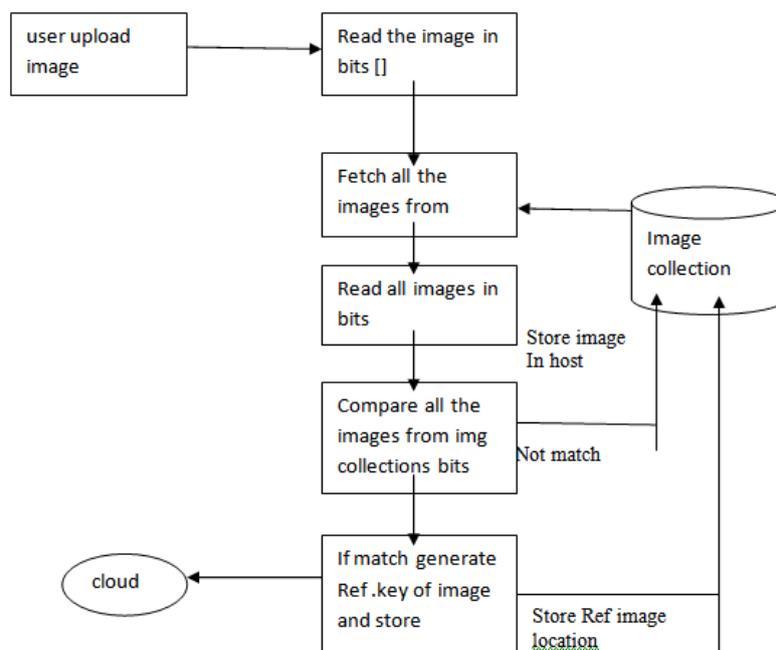
Figure 2: System architecture of proposed architecture
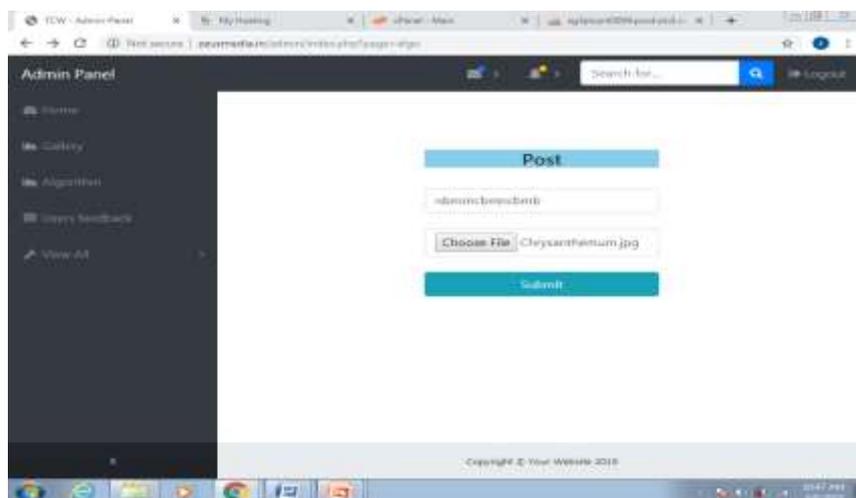
Figure 3: Screenshotof Uploading an image

Figure 4: Screenshot of comparison between images
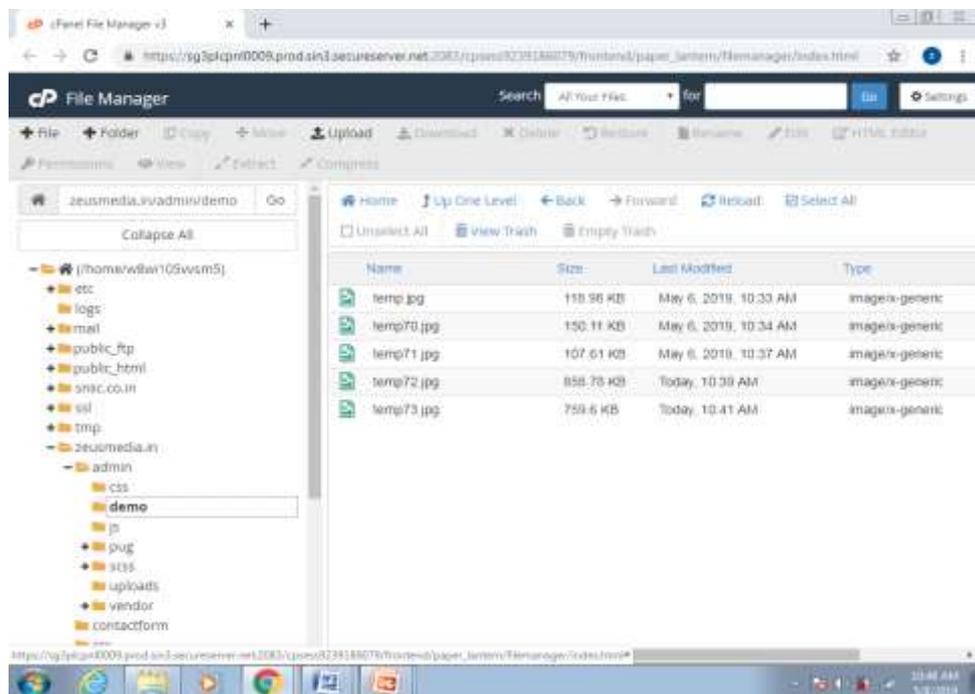
**Figure 5: Screenshot of Storing file**
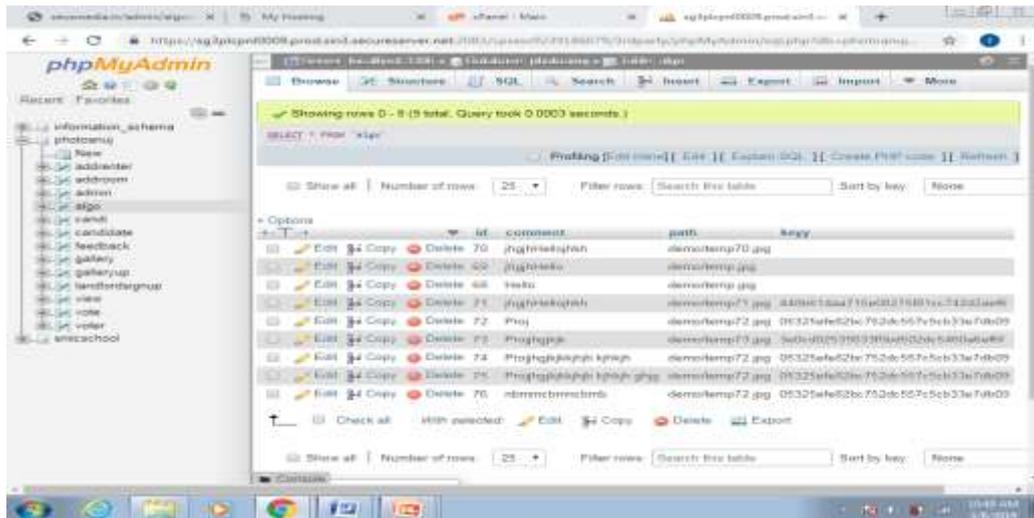
Figure 6: Screenshot of Generation of key

IV. PARAMETERS USED

It uses 3 parameters:

Size of the image: One of the important metadata of an image file is image size. The image size is determined on the basis of the width and height of an image. The image size should be smaller so that the comparison can easily be done, and file can be easily uploaded. Usually the size of the values is given in pixels.

Type of the image: The image should be converted into jpeg format before uploading so that comparison can be done easily. Then the jpeg image is converted into binary format and then between the images bit by bit comparison using string matching technique is done.

Zone: The image is compared from all the images uploaded on that zone, the zone can be any particular area i.e. town, city etc

V. RESULT ANALYSIS
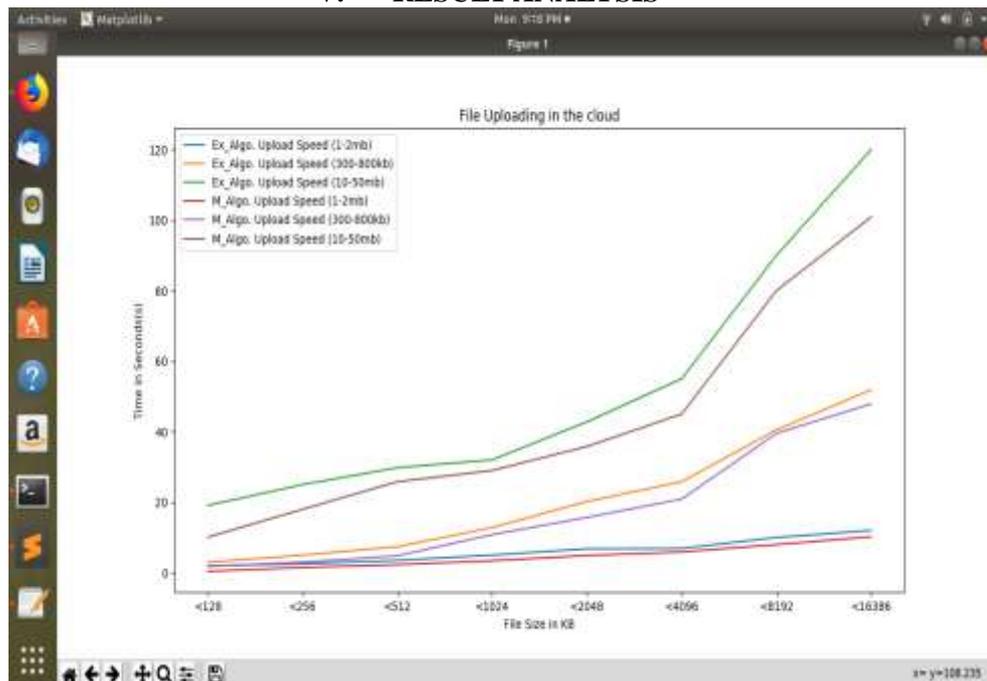
Figure 7: time taken by the existing approach and new novel approach to upload a file

In the above graph the comparison is done between the existing algorithm and the new algorithm. The comparison is done while uploading the image. The graph shows the time taken by all the algorithm for uploading Ex_Algo represents the existing algorithm and M_Algo represents the new algorithm. The outcome

of the graph shows that the time taken by the new algorithm is less than that of the existing algorithm for all the different size of the image.

VI. CONCLUSION

Our main objective is to focus on issues related to duplication and redundancy of data in cloud along with security and privacy aspect in cloud. The purpose of this work is to propose a new model MPODO which uses RSA algorithm for key generation and encryption. For data redundancy we use the data deduplication technique by compressing both image and comparing Discrete Cosine Transform (DCT) is applied to each block and then each block is subjected to quantization for compression and string-matching technique is used for comparison. The paper discusses the process of deduplication. The aim of proposed model is to maintain privacy and overcome deduplication overhead in cloud storage. In addition, it also addresses data availability issue and resists brute-force attacks.

REFERENCES

[1].   Rongzhi Wang (2017) "Research on data security technology based on cloud storage" in 13th Global Congress on Manufacturing and Management, GCMM 2016.
[2].   Zheng Yan, Mingjun Wang, Yuxiang Li and Athanasios V. Vasilakos, "Encrypted Data Management with Deduplication in Cloud Computing," IEEE Cloud Computing, March-April 2016, pp. 29-35.
[3].   Mohammed A. AlZain, Ben Soh and Eric Pardede "A New Approach Using Redundancy Technique to Improve Security in Cloud Computing" researchgate.net, publication,232733273
[4].   R. SHOBANA*, K. SHANTHA SHALINI, S. LEELAVATHY and V. SRIDEVI "DE-DUPLICATION OF DATA IN CLOUD" Int. J. Chem. Sci.: 14(4), 2016, 2933-2938ISSN 0972-768X
[5].   P.Sai Sandip1, P.Rajeshwari2, Dr.G.Vishnu Murthy3 (2017)" Utilization of Data Deduplication towards Improvement of Primary Storage Performance in Cloud" in International Journal of Innovative Research in Science, Engineering andTechnology.
[6].   Dhanaraj Suresh Patil, R. V. Mane, V.R.Ghorpade "Improving the Availability and Reducing Redundancy using Deduplication of Cloud Storage System" 978-1-5386-4008-1/17/$31.00 ©2017 IEEE
[7].   Mohammad Kamrul Islam and Ragib Hasan (2016) "Verifiable Data Redundancy in the Cloud" in 2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom) 978- 1-5090-3936-4/16 $31.00 © 2016 IEEE DOI 10.1109/BDCloud-SocialCom-SustainCom.2016.1629.
[8].   Khilan Goel and Ishant Goel (2016) "Cloud Computing based Social Media Model" in ISSA 2011: Information Security South Africa,(2016).
[9].   S KEERTHI 1*, MADHAVA REDDY A 2* "ENCRYPTED DATA MANAGEMENT WITH DEDUPLICATION IN CLOUD COMPUTING", et al, International Journal of Research Sciences and Advanced Engineering [IJRSAE]TM Volume 2, Issue 20, PP: 26 - 30, OCT - DEC' 2017.