

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 7, Issue 4, April 2018

## Deduplication on Encrypted Big data in Cloud

Sumedh Deshpande, Anupama Murkute, Saurabh Patil, Sachin Kamble, Prof. Swati Shekapure

Dept. of Computer Engineering, Marathwada Mitra Mandals College of Engineering, Pune, India

**ABSTRACT:** Cloud Computing provide services by rearranging different resources over internet. One of the service is datastorage. Data store in cloud is mainly in encrypted form. So traditional scheme has some weakness over security issue. So, in this paper to solve these issues with are taking help of hash code generation and double encryption techniques.

This helps us to Integrates with the cloud data deduplication with control access. Because of the authorized data holders who obtain the symmetric the encrypted data can also be securely accessed. Keys used for decryption of data. The results show the superior efficiency and effectiveness of the scheme for big data deduplication in cloud storage.

**KEYWORDS:** Cloud Computing, Data Encryption Security and Protection, Access control

### I. INTRODUCTION

The important and popular cloud service is data storage service. Cloud users upload personal or confidential content data to a cloud service provider (CSP) data center and allow to keep this data. In addition, the loss of control over Personal data leads to high data security risks, especially data privacy leaks. Due to the rapid Development of data mining and other analysis technologies, the question of privacy becomes serious. Therefore, a good practice is to encrypted data in order to guarantee the security of the data and the privacy of the user.

### II. LITERATURE SURVEY

Blowfish algorithm is use to provide encryption to the files Blowfish is a 64-bit block cipher with a variable-length key [1]

External adversaries which aim to extract secrete information As much as possible from public and private cloud [2]

Now, by taking advantage of the intra-user deduplication performed between clients and their DP, from a client point of view, the visibility of what is stored in the storage system is limited to his/her own account[3].

It is customary to define security in context like ours by using model involving game between two parties [4].

The advantage of deduplication unfortunately comes with a high cost in terms of new security and privacy challenges. Cloud deduplication is secure and efficient storage service which is as secure as block level deduplication [5]

### III. PROPOSED SYSTEM

#### ARCHITECTURE

In recent time, there are many problems of storage places in cloud. If data holder store file in cloud which is already available in cloud, so this is waste of memory. So, in this paper we remove the deduplication. Also, security issues are occur during the cloud server data storage. We are solving these problems with the help of double encryption technique and hash code techniques for generate hash value

# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 7, Issue 4, April 2018

## Data Holders

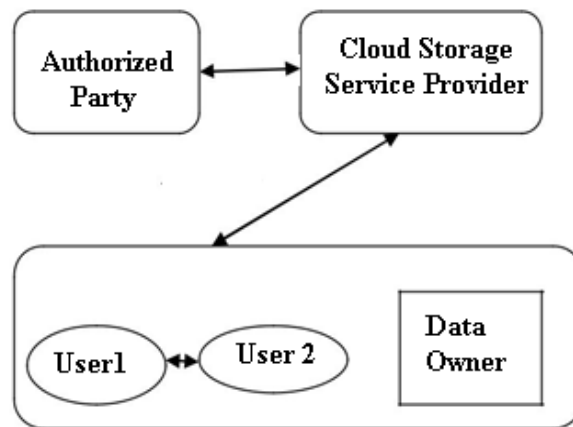


Fig 2: System architecture

**1.1 CSP (Cloud Service Provider):** The CSP (Cloud Service Provider) is allowing to data owner for storage services of data. It cannot be fully trusted. Therefore it is curious about the contents of stored data. It should perform honestly on data storage in order to gain profit.

**1.2 Data Holder:** The data holder can uploads and saves their data and files in the CSP. In this system is possible to number of data holders could save their files in encrypted raw data in the CSP. The file is regarded as data owner by the data holder that produces or creates theFile. The data holder is in normal form than the higher priority of owner.

**1.3 AP (Authorized Party):** An authorized party (AP) that is fully trusted by the data holders. The data holders to verify data ownership and handle data deduplication. It does not collude with CSP. In this case, CSP should not know the plain user data in its storage. AP cannot know the data stored in CSP.

## IV. ALGORITHMS

Step 1: Data owner Select File

Step 2: Encrypt File (For encryption we use BlowFish).

Step 3: Upload file on cloud or store the file on secondary storage.

Step 4: CSP or Controller check the duplicate file available on cloud or secondary storage.

Step 5: If found then remove the duplication and maintain index.

Step 6: On non-duplicate data CSP again check resemblance detection of similar chunk.

Step 7: If resemblance found then reduce it create delta store.

Step 8: Again on non-similar data check resemblance detection.

## International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

**Vol. 7, Issue 4, April 2018**

Step 9: If resemblance found then reduce it store similar data in delta store.

Step 10: Finally, non similar data stored into secondary storage or Cloud storage

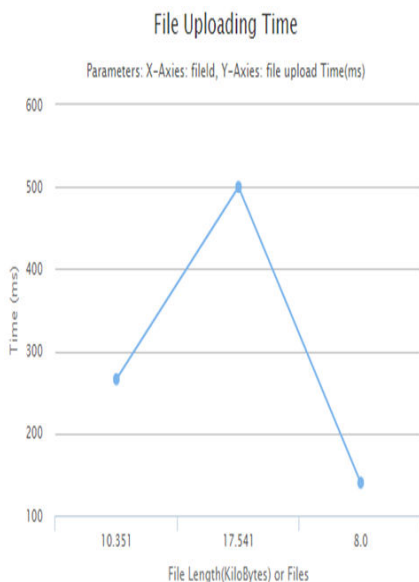
### V. SYSTEM ANALYSIS

In our experimental setup, In table 1, find out different file upload and time required for time for uploading that file. In our experimental setup, in our system first is uploading file size and time for that file.

Sr.No	File Size(Kb)	Time(ms)
1	10351	226
2	17541	500
3	8500	140

**Table1: File Uploading Time and Size**

From above data, In graph 1, we can see file size of 1 is 10351 kb is required time uploading is 226 ms, and file size of 2 is 1751 kb is required time uploading is 500ms



In our experimental setup, In table 2, find out different file download and time required for time for downloading that file. In our experimental setup, in our system first is uploading file size and time for that file and so on.

# International Journal of Innovative Research in Science, Engineering and Technology

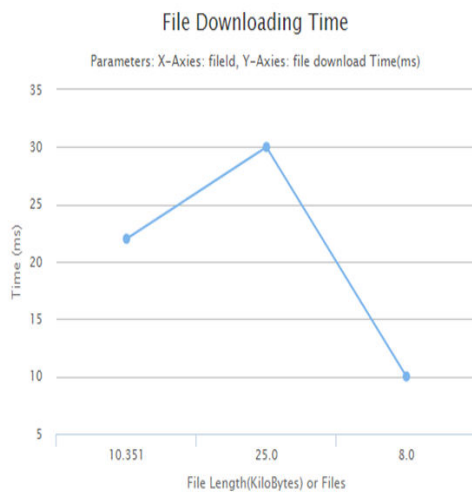
(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 7, Issue 4, April 2018

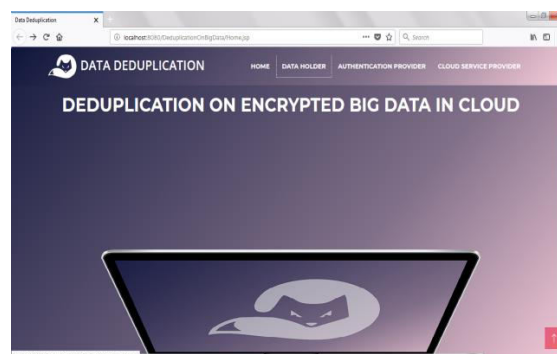
Sr.No	File Size(Kb)	Time(Sec)
1	10351	22
2	25000	30
3	8000	10

Table 2: File downloading Time and Size



From above data, In graph 2, we can see file size of 1 is 10351 kb is required time uploading is 22 second, and file size of 2 is 25000 kb is required time uploading is 30 sec. and so on

## VI. RESULT



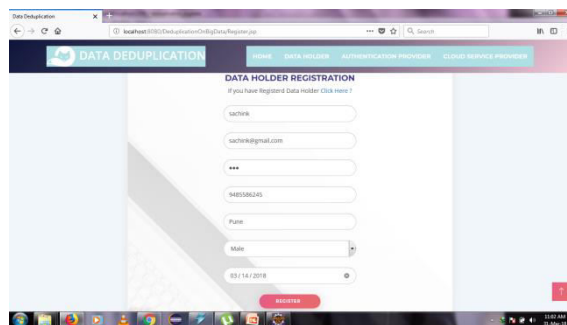
# International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

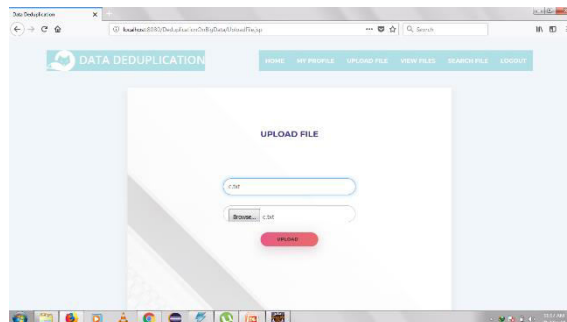
Visit: [www.ijirset.com](http://www.ijirset.com)

Vol. 7, Issue 4, April 2018

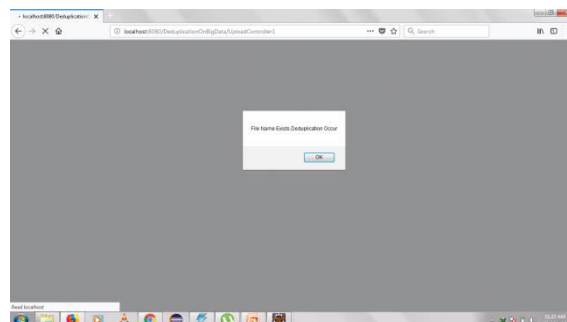
## Home Page



## Registration Page



## File Uploading



## VII. CONCLUSION

Discuss about to managing encrypted data with deduplication is for achieving a successful cloud storage service, for big data storage. In this paper, proposed a scheme to manage the encrypted big data in cloud with deduplication based on PRE and ownership challenge. This scheme can flexibly support data update and sharing with deduplication even when the data holders are offline.

# International Journal of Innovative Research in Science, Engineering and Technology

*(A High Impact Factor, Monthly, Peer Reviewed Journal)*

Visit: [www.ijirset.com](http://www.ijirset.com)

**Vol. 7, Issue 4, April 2018**

## REFERENCES

- [1] MsNehaKhatri – Valmik1, Prof. V. K Kshirsagar2 proposed paper on Blowfish algorithm.
- [2] Chi Chen at. Al. proposed An Efficient Privacy-Preserving Ranked Keyword Search Method IEEE 2016.
- [3] Chunhua Su at. al. proposed Analysis and Improvement of Privacy-Preserving Frequent Item Protocol for Accountable Computation Framework IEEE 2012.
- [4] Cheng Huang and Rongxing Lu proposed EFPA: Efficient and Flexible Privacy-Preserving Mining of Association Rule in Cloud in IEEE 2015.
- [5] Bharath K. Samanthula at. Al. k-Nearest Neighbor Classification over Semantically Secure Encrypted Relational Data MAY 2015.
- [6] Lichun Li at. al. Privacy-Preserving-Outsourced Association Rule Mining on Vertically Partitioned Databases in AUGUST 2016.
- [7] M. Bellare, S. Keelveedhi, and T. Ristenpart,
- [8] “DupLESS: Serveraided encryption for deduplicated storage,” in Proc. 22nd USENIXConf. Secur., 2013, pp. 179–194.