International Journal of Marketing Management

ISSN 2454 - 5007

www.ijmm.net

Email ID: editor@ijmm.net , ijmm.editor9@gmail.com

International Journal of Marketing Management

ISSN 2454-5007, www.ijmm.net

Vol. 10, Issuse.2, May 2022

Multi-Chunk Data De duplication among Multi-Cloud with Multi-Key User Level Security

1B. Neeharika, 2Ch.Sarada

Abstract: In the cloud, When data is de-duplicated, it becomes a single piece of information. By using the Cloud Data De-Duplication approach, users will be required to upload only one copy of each file, resulting in significant storage and performance gains. As it is, the existing system has several flaws. For starters, there is a lack of user-level protection and the level of security given by the Cloud is lower because the entire data file resides in a single Cloud. This paper proposes a new system to address these issues. Using the new technology, data is uploaded to multi-cloud storage in a nonduplicate encrypted format. There are no two data partitions stored on the same cloud. Cloud data security is now better protected as a result of these changes. The new system generates a One Time Password (OTP) for every user login in order to further enhance cloud data security. A pre-generated hash code is used to schedule the de-duplication process. DriveHQ's experimental analysis runs on Java/Jsp and MySQL.

Keywords- Cloud Storage, De-Deduplication, Pre-Encryption, and Scheduling An ABE, Asymmetric Key, OTP, and CSP (Cryptographic Signature Protocol) are all examples of de-dedupe.

1. INTRODUCTION

The Cloud Service Provider supplies all the hardware, software, and operating systems that are needed to run a business. The Cloud System is a pay-per-use API and centralized data storage with all the latest security and privacy features. The cloud is one of the most adaptable and user-friendly methods of storing and retrieving data on the Internet today. Large amounts of data can be stored on the cloud without having to increase storage, update the devices and their related software's anywhere in the Internet's virtual worlds.Customers can choose from a wide range of services provided by Cloud Service Providers (CSPs). Using the Cloud Storage API, web applications, web services, and other web-centric applications can access the cloud issues like duplication and security enhancement that need to be addressed.

Cloud Users will continue to upload and download files on a second-to-second basis. Duplication occurs when numerous users upload the same data or when a single user uploads the same file with the same content more than once. Duplicate data wastes storage space, increases maintenance costs, and poses a security risk for businesses.Data sensitivity necessitates the security of sensitive personal information that is saved for distribution. Businesses, Consumers, Medical, Retails like Amazon and FlipCart, Scientific & Research, and many more use digital storage in a safe encrypted way [2]. Multi-cloud Recently, a de-duplication strategy was implemented to address the issue of non-duplicate data being uploaded to many cloud services. Thus, not only is a single Cloud subject to

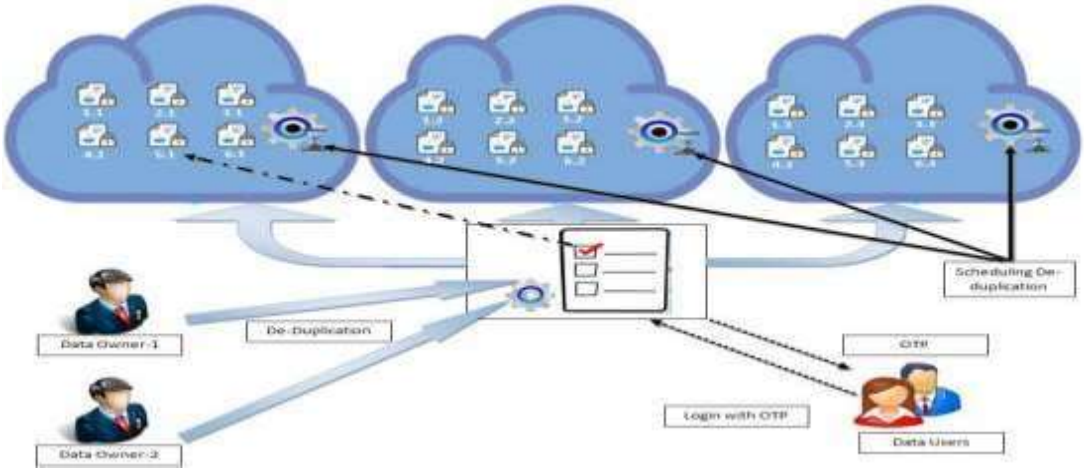
1,2 Department of Computer Science and Engineering, CVR College of Engineering Hyderabad, India
neeharikabehara@gmail.com, sharada.ch@gmail.com

International Journal of Marketing Management

ISSN 2454-5007, www.ijmm.net

Vol. 10, Issuse.2, May 2022

a duplication check, but several Clouds as well. It reduces storage costs while also requiring minimal upkeep. Data encryption is used to ensure the security and confidentiality of sensitive information in this system. When it comes to reducing storage waste, keeping costs down, and enhancing performance in the Cloud, a De-duplication approach can



be a lifesaver. Nevertheless, this solution does not provide adequate protection for data; all of the files are uploaded to a single cloud. As a result, there is a significant likelihood of a data breach. In addition, there is no user-level security in the current system. This research proposes an enhanced de-duplication technique to enhance the current system.In this paper, data De-duplication technique is proposed, that stores the Encrypted chunks of a data file onmultiple Cloud platform where each chunk is placed at different cloud. This feature increases security to the dataas the hacker cannot know where which data block is stored. This proposed solution also provides user levelsecuritybyprovidingOTPtothedataowners /datausersforeverylogin.TheFig.1showsthepro posedsystemmodel.

Figure1.ProposedSystemModel

De-duplication techniques are also discussed in this study as a means to alleviate the burden of data administration and search. HashCode is calculated based on a file's data and compared to other HashCodes to ensure that no two files have the same HashCode.

2. RELATEDWORK

De-duplication is conducted on various Clouds in an existing work [1] heterogeneous data storage management approach. Storage waste is decreased because of the elimination of redundant files in many Clouds. De-duplication is currently done by either the data owner or a trusted third party under the current system.Data is encrypted and kept in the cloud, which inhibits illegal access, according to previous studies [4-6]. The encrypted data can only be decrypted by those who are allowed to do so.. Access control to the encrypted cloud data was made possible by the use of Attribute-Based encryption (ABE) [7-10]. Data is encrypted according to the access hierarchy defined by these parameters, which are used to identify and authenticate users.

The encrypted data can only be decrypted by the users who meet the access structure's requirements. Key-Policy ABE (KP-ABE) and Cipher-Policy ABE (CP-ABE) are two subcategories of ABE.In the current system, data access control is resorted based on the identification of the user during the de-duplication process by applying ABE [7-10]. As

ISSN 2454-5007, www.ijmm.net
Vol. 10, Issuse.2, May 2022

a result of ABE, data owners can issue encryption keys to data users only when they have access rights. This is an important benefit of ABE. The decryption keys that are supplied to the data consumers are unique and cannot be used by anyone else. This is done using the CP-ABE algorithm.
ABE decryption keys are generated for each user based on their attribute ID, which is known as a secret key attribute in this approach. Not only is de-duplication done on numerous clouds, but it also tends to keep

divided into DEK1 and DEK2. The public key of the authorized party (AP) is used to encrypt DEK1, whereas the public key of the user is used to encrypt DEK2. Encrypted DEK1 and DEK2 are sent to the CSP together with encrypted data once the encryption process is complete.Thedisadvantagewiththeexistingsystemisthatifthesymmetrickeyisknowntothehackers,theycaneasilydecrypt the data and can access it. To overcome this disadvantage, we proposed a data De-duplication techniquewhich divides the file into blocks and then each block is encrypted separately, and each block is stored indifferentclouds.

As a result, the cloud storage of sensitive personal information requires a heavy blanket of security to protect it. However, the current technology does not provide the level of security that is necessary. This difficulty has been attributed to two factors. Because the entire uploaded file is stored in a single cloud, this is one of the reasons. The information contained in a single file is not dispersed over multiple clouds. In other words, if the symmetric key used to encrypt the entire file is known to the hackers, they will be able to easily break into the data. The second reason has to do with granting users' access to the Cloud Service Provider when they upload or download data from it. Authorized access is supported by the current system; however, the authorization key is not changed on a regular basis. As a result, an intruder will find it much easier to launch an attack. A new mechanism has been put in place to enhance the safety of user data.

unique data across all of the clouds as a result of the de-duplication. Uploading a file generates a HashCode for the data, which is then used to check for duplicates. In order to verify the file's authenticity, a signed HashCode is generated after HashCode. In order to verify the authenticity of the signed HashCode, it is sent to the CSP. The data owner encrypts the data with a randomly generated symmetric key data encryption key if no CSPs have the same data (DEK). After then, DEK is

Blocks/chunks are used to divide the data across many clouds, with each block/chunk being stored at a distinct cloud, according to the proposed approach. Each block of data is encrypted separately using the new system's encryption keys. An invader would have their work cut out for them if they tried to enter through this method. The One Time Password (OTP) feature of the new system further enhances data security.The rest of the paper follows the same structure. Section 3 explains the preliminaries and definitions utilized in our plan. Section 4 details the proposed scheme's design, followed by Section 5's security analysis. Section 6 discusses the results of an experimental study. Section 7 brings the paper to a close and lays out the paper's future directions.

3. PRELIMINARIESANDDEFINITIONS

Thenotationswhichareusedinthispaperareliste dbelow.

A. ABE:AttributeBasedEncryption
B. KGC:KeyGenerationCenter
C. AP:AuthorizedParty
D. Pu1Key:PublicKey1
E. Pu2Key:PublicKey2
F. Pr1Key:PrivateKey1
G. Pr2Key:PrivateKey2
H. Sk1:SecretKey1
I. Sk2:SecretKey2
J. Sk3:SymmetricKey3

K. Sk3(a):SymmetricKeySplit1
L. Sk3(b):SymmetricKeySplit2
M. PRE:PreEncryptionPassword
N. Data-CT:CipherTextofdata

International Journal of Marketing Management
ISSN 2454-5007, www.ijmm.net
Vol. 10, Issuse.2, May 2022

O. f:file
P. fd:fileData
Q. hC:HashCodeofthefd(fileData)
R. fn:fileName
S. f1,f2….:otherfilesstored
T. n:noofcloudsselectedforstorageandthenfilesblockstobegenerated
U. DU:DataUsers
V. DO:DataOwners
W. CSP: Cloud Service Provider
X. OTP:OneTimePassword
Y. DEK:dataencryionkey
Z. DEK1:dataencryptionkey1AA.DEK2:dataencryptionkey2 AB.AES:AdvancedEncryptionStandardAC.ABE: AttributeBasedEncryption

4. OVERVIEWOFPROPOSEDSYSTEM

De-duplication in a multi-cloud context is made possible with the suggested method since it ensures complete data security. The proposed system's summarized model can be seen in Figure 2.

4.1 Model for Summerized Flow

A simplified flow diagram of the proposed system is depicted in Figure 2. Files are broken up into blocks depending on the number of cloud storage options accessible to the data owner/user when uploading. That is to say, the number of blocks is equal to the data/cloud count. In each cloud, the blocks and encryption keys are stored independently. At the moment of downloading, each block is downloaded, decrypted, and concatenated to form the original data file.

International Journal of Marketing Management

ISSN 2454-5007, www.ijmm.net

Vol. 10, Issuse.2, May 2022

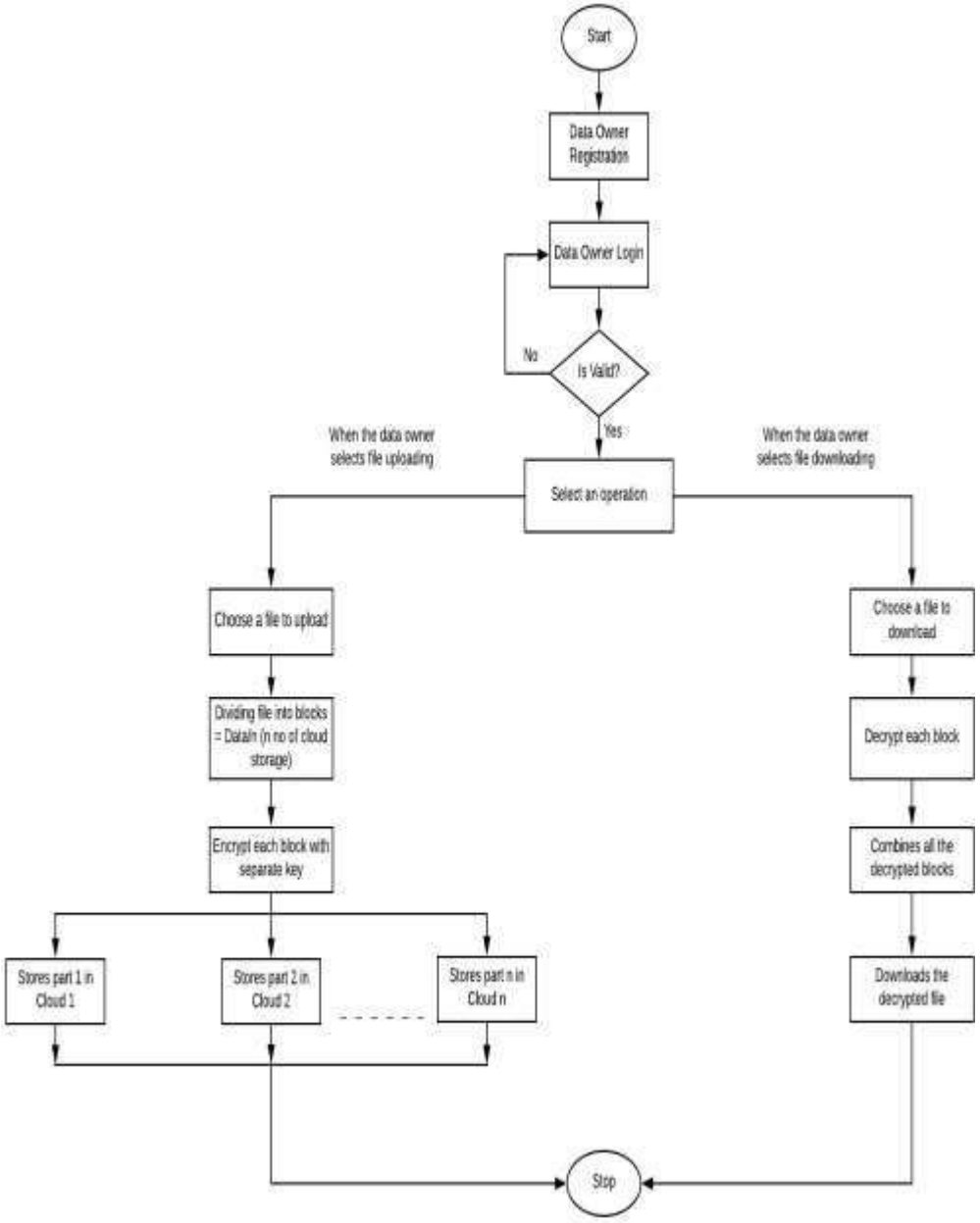


Figure2.SummerizedFlowModel

4.1 UploadProcess

Figure 3 depicts the upload of a file in great detail. To upload or download a file, the data owner must first log in. An OTP is created and delivered to the data owner's registered email address each time he or she logs in. The user can access their data after they enter the OTP generated by the system.Before a file may be uploaded, it is checked to see if it has already been stored somewhere else. An algorithm is used to determine whether or not two files are identical. The link to an existing file is displayed if the currently created HashCode matches the previously generated HashCode. Uploading begins when the HashCode does not match the previously uploaded HashCode..

International Journal of Marketing Management

ISSN 2454-5007, www.ijmm.net

Vol. 10, Issuse.2, May 2022

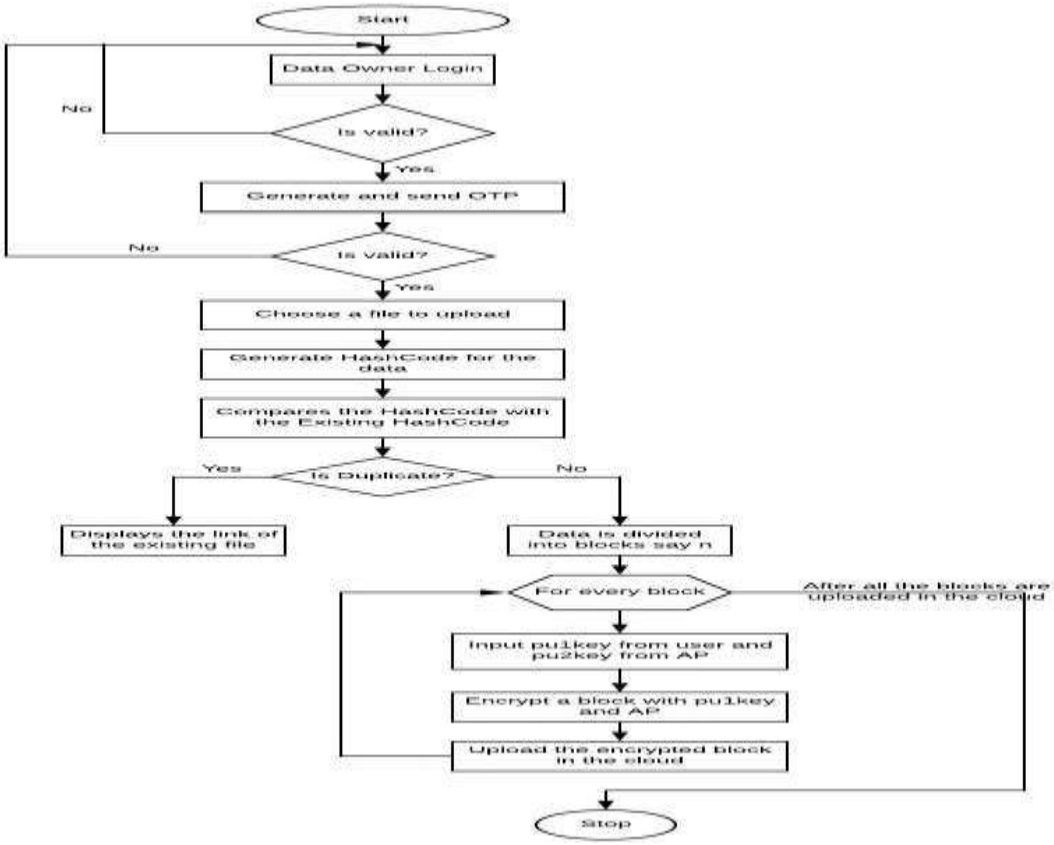


Figure4.Processofdownloadingafile

An example of how to download a file is shown in Figure 4. A one-time password (OTP) is generated and sent to the subscribed email address of the user before the file may be downloaded. Once the user has successfully logged in, he can download the file for which he needs the data owner's permission. When a user requests access to a file, he or she must first get a secret key for each block of encrypted information from both the data owner and an authorized person (AP). The requested file blocks are decrypted and then concatenated to form the file contents if the user-entered keys are genuine.

SchedulingDe-Duplication:

Admins can also use this paper's automated scheduling feature. This is applied to any duplicate files that already exist. The HashCode of all files is compared when setting a job. If the identical HashCode is found, then it merges the duplicate copies into a single file by deleting all of the duplicates..

SECURITYANALYSIS

OutputScreen1:Originalcontentofafile

Figure5.Originalcontentofafile

DSA, KP-ABE, and AES Secret Key algorithms have been implemented in this paper. Access Policy can be effectively addressed by using Key-Policy Attribute Based Encryption (KP- ABE). In order to protect passwords and generate certificates for authentication, the DSA algorithm is utilized. File data is encrypted and decrypted using the KP-ABE method. Symmetric keys are generated using the AES method. AES's symmetric key is generated through the use of random number generation. HashCode generation and signed HashCode generation employ the SHA algorithm. Data is verified using HashCode.The suggested scheme's security is enhanced by integrating all of the above-mentioned qualities, thereby ensuring privacy, reliability, and confidence for the data owner or user.

EXPERIMENTALSTUDY

Theproposedsystemistestedonmulticloudenvironmentwiththreeclouds.WetestedthisinDriveHQcloudenvironment.

International Journal of Marketing Management

ISSN 2454-5007, www.ijmm.net

Vol. 10, Issuse.2, May 2022

This screen shows the content of the file which is to be uploaded. This screen appears as soon as the file isselectedforuploading.

OutputScreen2:HashCodegenerationofaselectedfile

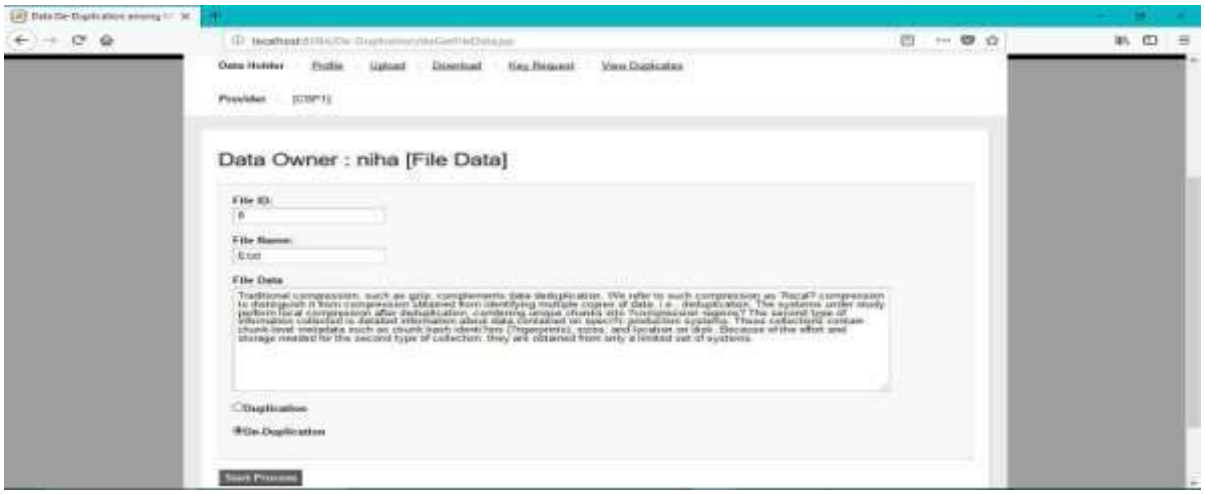


Figure 6.HashCode generationof aselected file

When uploading a file, a HashCode is created based on the file's content. In order to determine if a file is already in the Cloud, this HashCode is needed. If the file doesn't already exist, it's broken up into blocks, encrypted, and sent to the cloud as a backup. The data is encrypted one block at a time. We do this to ensure that the blocks are uploaded to many cloud providers to increase the data's security.

OutputScreen3:Encryptionoftheblocks

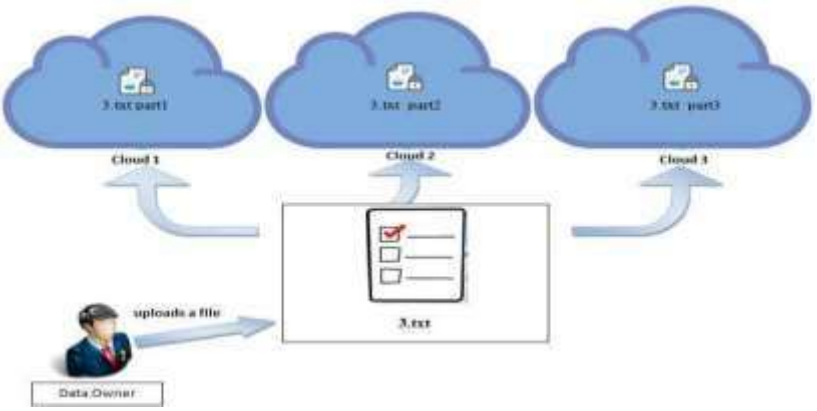


Figure7.Filestoredinblocks

International Journal of Marketing Management

ISSN 2454-5007, www.ijmm.net

Vol. 10, Issuse.2, May 2022

Files are being uploaded to three different Cloud storage services, as shown in this screenshot. A file called 3.txt was picked for uploading, and the next panels reveal the encrypted form of all three partitions in the cloud.

5. CONCLUSIONANDFUTUREWORK

Data is encrypted, partitioned, and stored across many clouds in the proposed system. Each partition has its own unique encryption key. This multi-key encryption enhances data protection. This system makes use of the One Time Password approach to enhance data security.

Using solely the AES encryption method, the suggested system creates all of its keys. In the future, we plan to work on a variety of encryption methods to further enhance data security. Media assets, such as audio, video, and photos, can also benefit from De-Duplication.

REFERENCES

[1] Deduplication in Cloud Computing for heterogeneous data storage management Lifang Zhang, Wenxiu Ding, and Qinghua Zheng are all IEEE members, as is Zheng Yan, a senior member. [2] Authenticated Data Dedupe This is the name of Mark Storer He is Kevin Greenan. Darrell Long-term storage systems research at the Ethan L. Miller Storage Systems Research Center at the University of California, Santa Cruz StorageSS'08 was held in Fairfax, Virginia, on October 31, 2008.

[3]A Cloud Computing Study on Approved Deduplication Techniques Computer Engineering & Technology (IJARCET) Volume 3, Issue 12, April 2014, by Bhushan Choudhary and Amit Dravid

Proc. 2009 ACM Workshop Cloud Comput.Secur.: "Controlling data in the cloud: outsourcing compute without outsourcing control," in Proceedings of the 2009 ACM Workshop on Cloud Computing Security, pp. 85-90, 2009.

Cryptographic cloud storage, Springer, 2010, S. Kamara and K. Lauter, pp. 136-149 in: Financial Cryptography and Data Security.

"Efficient information retrieval for ranked queries in cost-effective cloud environments" was published in Proc. 2012 IEEE INFOCOM, pp. 2581-2585.

A. Sahai and B Waters, "Ciphertext-Policy Attribute-Based Encryption," in IEEE Symp. Secur. Privacy (SP'07), pp. 321-334, 2007, are the authors of this paper.

[8] There is an article by V. Goyal et al. titled "Attribute-based encryption for fine-grained access control of encrypted data," published in the proceedings of the 13th ACM Conference on Computer Communications Security (PCCS), in which they describe their approach.

Proc. of the 11th Annual International Conference on Information Security and Cryptology, pp. 20–36, 2008. [9] S. Muller, S. Katzen beisser, and C. Eckert "Distributed attribute-based encryption"

(10) "Fuzzy identity-based encryption," A. Sahai and B. Waters, in Proceedings of the 24th International Conference on Theory and Application of Cryptographic Technology, pp. 457–473, 2005.