Deduplication for Cloud based on AES Algorithm

R. Naresh, Vibhor Lohani, Naman Agarwal

Abstract**:** ***This paper provides and analyses a new scheme to address the issue of everlasting cloud storage access systems by providing AES Algorithm based solution. Addressing the issue of Data storage in the recent times is the aim of the paper and is being done using the recent AES Algorithm. Data security and time constraint for data retrival from the cloud server is also taken into account while implementing the scheme***

Index Terms**:** ***De duplication , Storage , Route key , AES Algorithm***

I. INTRODUCTION

Data is the key factor in the modern era. Starting from Food Items, Groceries and till the high end satellite and rocket mechanisms , datas play an important role. A separate study is being implemented in Grade Schools and Institutions on data analysis so as to cater to the evergrowing era of data science and its applications. More than 2.5*10^18 bytes of information are produced consistently. More than 90% of the above information has been made just in the recent years alone. This number will arrive at 35 ZB in 2025, which has demonstrated to be too bigger than imagination and control.

Data storage mediums ranged during the earlier days were Selectron tube , Punched card , Punched tapes , Drum memory and then the IBM HDD.. Depending on the type of data storage medium, the data storage type also changed and this digital era , we store data in the server with the specialized keys for security concerns. Use of cloud computing is increasing. Cloud optimization is increasing. Effective use of cloud resources is the want of this time , as redundant datas are stored in the cloud again and again. This causes inefficient data storage in cloud and also affects the upload bandwidth. Data security is the major criteria while accessing data from and to the server. We need to reduce the load on the server or cloud storage so as to make it free and perform hassle free data transactions. Removal of duplicate data from the cloud and provide and access to the files will be the prime address issue of this paper implementation. The paper aims to free space, bandwidth and storage in cloud. The suggested approach is to remove the redundant data , where every user has been assigned some access according to the duplication check & each user have their prioroty token. Hybrid cloud organization is deployed to accomplish the deduplication in cloud. The suggested method is extra secure and deplete less resources of the cloud. Also it is shown that the proposed scheme has very less overhead in removal of duplication as compared to the ordinary technique employed for deduplication . In this paper dedulication is done using the AES Algorithm.

II. RELATED WORK

The strategies of de duplication can be of two types. file-level and block-level de duplication. The file-level de duplication removes the duplicate data copy at the file level if two files have the same value and are identical .This File level de duplication presumes low head and the efficiency is low. Incase of block-level deduplication every single input file is divided in multiple blocks of presumably fixed or variable size and then use hash value of each block to remove the block that is already stored in cloud.

Though the cloud infrastructures are having a turbid and heist infrastructure and are also more safety secured comparing to the personal systems, they also do face security threats for data integrity and security. Encryption alone will not be sufficient to solve the problem of cloud security. Data leakage from the cloud will be still possible due to the decryption key availability. More related works has been done inorder to provide cloud security while data transfer and to enhance the optimal usage of cloud storage.

But still problem of data security persists with the want to access the cloud for data retrival. De-duplication systems so far are available only for single-server setting. In current viabilities , let check the option of usage of the cloud storage . Incase if User1 uploads a file x to the cloud , decipers it using his own algorithm and saves it to the cloud. User2 also has the same file , but decipers it using another algorithms and saves it to the cloud with his key. The cloud is unable to access the file as it is being stored as encrypted text and hence the same file is being stored in the cloud under two different names and with two different text deciphers .

This causes affecting cloud storage facilities and optimal usage of the cloud storage will not be acheieved. Hence they require that the de-duplication storage systems provide reliability comparable to other higher end systems.. Since many of the systems for deduplication and systems for storage, are designed by users for user applications for higher accuracy, and in special goal limits where the data stored is extremely important and should be preserved for more longer time duration cost increases for content strorage as well as for the keys storage. Moreover there will be a higher increase bandwidth with along with the upload time. The Bandwidth consumed is directly proportional to the upload time . It increases are the time increases and vice versa. This will cause a extra overhead on the cloud and the costs pertaining to the bandwidth and the cloud storage infrastructure may also increase.

Revised Manuscript Received on May 30, 2020.
***** Correspondence Author

Dr. R.Naresh* , Associate Professor in the Department of Computer Science and Engineering at SRM Institute of Science and Technology.

Vibhor Lohani , Student. Department of Computer Science and Engineering at SRM Institute of Science and Technology.

Naman Agarwal , Student. Department of Computer Science and Engineering at SRM Institute of Science and Technology.

III. METHODOLOGY

We propose a quality-based stockpiling structure which employs encryption which are based on AES Algorithms that is a rising encryption innovation to address difficulties of secure information sharing. AES based framework plan has been actualized for fine grained texts or images get to control of archives.

The data or reports are encoded with an route key and can be decoded. The key ensures that the shared file is correctly stored and the address is recorded. Earlier advancements show the capacity of proposed algorithm in giving constant encryption and decoding administrations on changing the number of characteristics, record sizes and types. To empower the de-duplication and conveyed stockpiling of the information across the storage , we utilize two route cloud in our cloud framework. A private cloud controls the calaculation and an open cloud deals with the capacity. The private cloud is furnished with a route key related with the comparing ciphertext, which it can alter the ciphertext for more than one access arrangement into cipher texts of the equivalent plaintext under some other access approaches without monitoring the hidden plaintext.

When a private cloud receives a request for storage it immediately checks the potency of the uploaded item through the proof attached. If the proof proves to be potent, the private cloud uses algorithm for tag match to determine that same data underlying the ciphertext is stored. It reconstructs the ciphertext into a ciphertext belonging to the same plaintext which is the union of both access policy like public and private cloud. Proof ownership of the file is used to achieve the approach of deduplication securely and efficiently. Storage system uses ciphertext encryption policy based on attribute and also supports secure de-duplication.

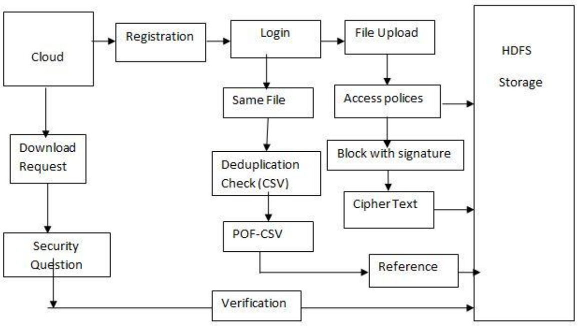


Fig .System process flow diagram

The cloud user has been registered and login to the cloud storage. The file is uploaded and tagged by using MD5, keys are generated using SHA-256 and stored in HDFS storage. The file undergoes duplication check(CSV) with the original file. If the file is matched with original file the download request has been given followed by POF- verification then the file is accessed from the HDFS storage. During the duplication check existing data can be eliminated using Delete option.

IV. SYSTEM IMPLEMENTATION

The proposed system implementation consists of four different modules as the following , cloud user registration ,File upload and access policies ,Dededuplication method & Download user files

A. Cloud User Registration

The cloud user will first fill the user details such as his/her Name and password, email, access policies, mobile, Date of birth. He / She can then use these to login into the system with username, password. Once user name and password is proved to be potent, the user profile screen will be displayed.

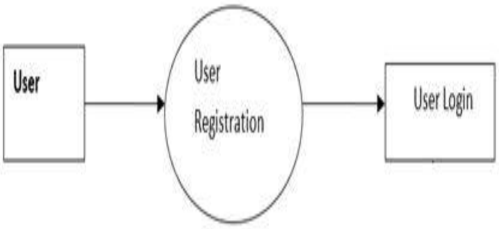


Fig .Cloud user registration form

B. File Upload with Access Policies

The registered user can then have the privilege to access and upload any file or document to the cloud . The upload file is splitted , each splitted file is tagged and a spontaneous product key is generated and is stored in the cloud with reference to the file that has been uploaded. The UFile upload registration form block is as follows

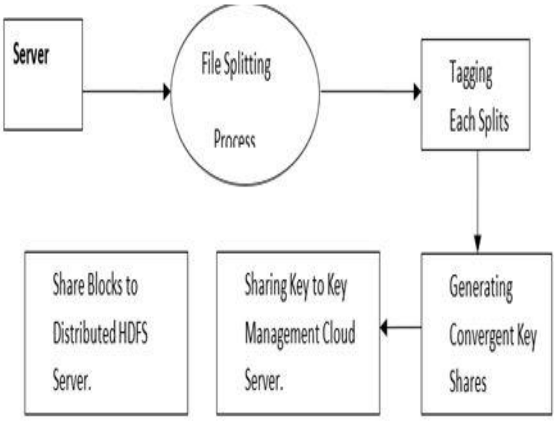

Server
File Splitting Process
Tagging Each Splits
Share Blocks to Distributed HDFS Server.
Sharing Key to Key Management Cloud Server.
Generating Convergent Key Shares

Fig .File upload form

Encryption of blocks by AES algorithm is known as asymmetric cryptography algorithm. It really means that it works on two various keys i.e. Public and Private Key. As we can understand from its name, Public Key is given to everybody and Private key is private to individuals. The plain text is encrypted to cipher text and stored in slave system. Distributed HDFS Providers are used for storing the blocks.

C. File level Deduplication

Data deduplication on file-level contrasts a file that is to be backed or archive, with the copies that are already there in the cloud. This is achieved by checking its attributes against index. If the file is novel, it will be stored and the index is refurbished and only a pointer to existing file is stored. The resulting situation consists only one instance of the file and subsequent copies are recouped with a reference to the original file.

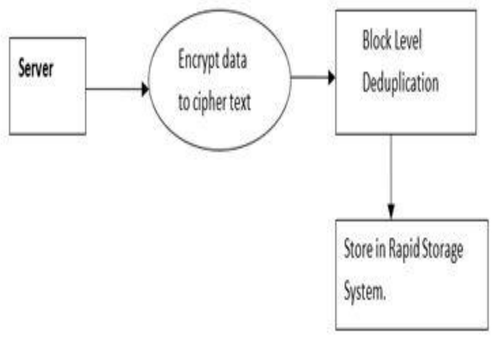

Server
Encrypt data to cipher text
Block Level Deduplication
Store in Rapid Storage System.

Fig. Detection Deduplication Application

Another signature match scrutinizing technique looks within a file and saves very unique repetitions of each block. All blocks are divided into dollops with the same fixed length. Each dollop of data is processed using a AES Algorithm.

D. User File Download Module

In the final model user is requested to download their document which they had uploaded in HDFS storage. In this download demand will analysis the attribute of the user and once it is verified it will ask security questions for that particular file. After the process completion it needs proper verification of the ownership

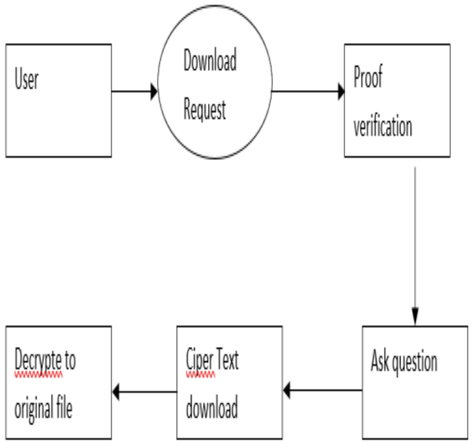


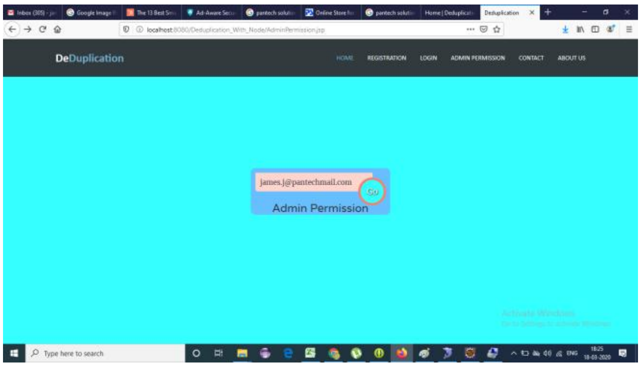
Fig. User File Download Module

Afterwards the original content is successfully verified, it is decrypted by requesting the Distributed HDFS storage where HDFS storage request key management slave for keys to decrypt and finally the original content is received by the user.

V.SCREENSHOTS OF IMPLEMENTATION

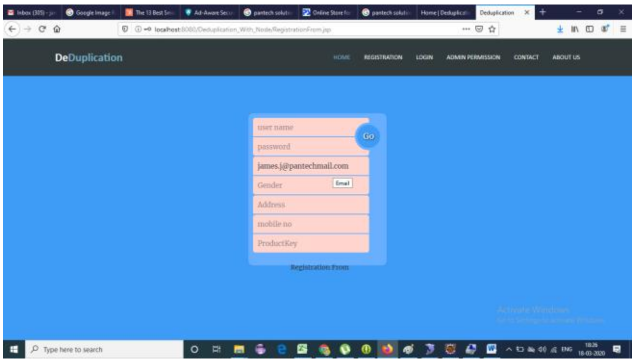
The process of the deduplication has been successfully implemented and the following screenshots of the paper has been taken for consideration.

A user name James and Naman has been created in the cloud and their details and user credentials has been saved in the SQL Server as per the format. Credentials are saved and cannot be accessed by any other users and is saved coherently.



james.j@pantechmail.com

Fig . User Registration Screenshot



james.j@pantechmail.com

Fig . User Registration Attributes

The above screenshots presents that multiple user credentials that has been saved in the server to allow ease of access to the user and to eradicate or completely remove the fake users from logging in or trying to download the data stored in the cloud server.

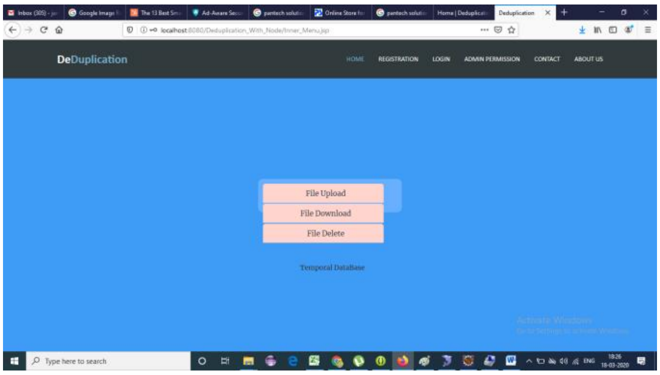
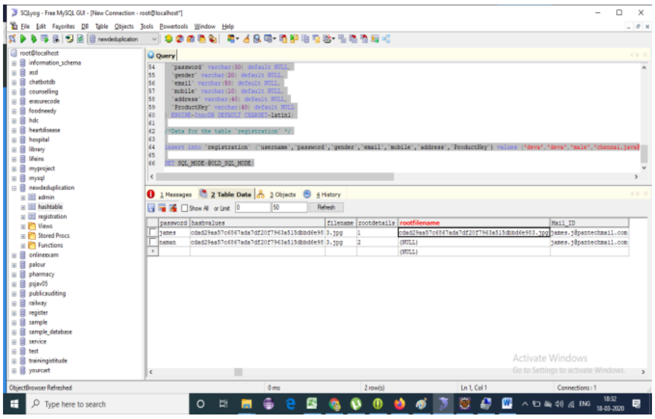


Fig . User Registration Attributes

The registered user can upload, Download and Delete the files that he has contributed. This allows the user to have access to the cloud environment . The uploaded file or document is stored in the cloud using the AES Algorithm , so that the file has been encrypted and has been assigned a new or unique product key / route key at the time of storage in the cloud. The unique product key is shown here for reference.(Fig below)Another second user named Naman has been registered. Incase User 2 ,Naman has accessed the cloud and is trying to save the same file that has been uploaded by user 1 James as in the implementation ,there will be a notification that , the file has been uploaded successfully in the cloud , but it is not actually stored in the cloud nor an product key assigned. The this file, thus saving the cloud storage.



```
  `password` varchar(30) default NULL,
  `gender` varchar(20) default NULL,
  `email` varchar(50) default NULL,
  `mobile` varchar(10) default NULL,
  `address` varchar(40) default NULL,
  `ProductKey` varchar(40) default NULL
) ENGINE=InnoDB DEFAULT CHARSET=latin1;

/*Data for the table `registration` */

insert into `registration` (`username`,`password`,`gender`,`email`,`mobile`,`address`,`ProductKey`) values ('deva','deva','male','chennai.java@

SET SQL_MODE=@OLD_SQL_MODE;
```

password	hashvalues	filename	rootdetails	rootfilename	Mail_ID
james	cdad29aa57c6867ada7df20f7963a515dbbd6e98	3.jpg	1	cdad29aa57c6867ada7df20f7963a515dbbd6e983.jpg	james.j@pantechmail.com
naman	cdad29aa57c6867ada7df20f7963a515dbbd6e98	3.jpg	2	(NULL)	james.j@pantechmail.com
				(NULL)	

Fig . User Registration Attributes

Both the users can access the original file , they can download it and have private access.. But only the original file has been stored in the cloud and only the original file has been assigned a product or route key. As the file is stored in the encrypted format , that users having the privilege to access only can access the files under their domain.

VI. RESULTS AND DISCUSSIONS

The system successfully adds users and admins and provides a platform to upload the files. The files are being encrypted successfully to avoid any unauthorized interference and modification using the key provided by the user at the time of signing up.

VII. CONCLUSION & FUTURE ENHANCEMENT

Data de-duplication merchants are providing their products with enhanced capabilities which happen to increase its rate of adoption in and overthrow the challenges from substitutes such as cloud, tape and sometimes regular disks in a server. The systems are becoming more mature from disk storage systems involving de-duplication to genuinely complete data protection devices that can be integrated into applications and backup softwares for upgraded management and efficiency.

REFERENCES

1. B. Martini, K. R. Choo and D. Immediate Cloud Storage Forensics. Syngress Publishing/Elsevier,2014. quick/978-0-12-419970-5
2. Z. Yan, M. Wang, Y. Li and A. V. Vasilakos, "Encrypted Data Management with Deduplication in Cloud Computing," in *IEEE Cloud Computing*, vol. 3, no. 2, pp. 28-35, Mar.-Apr. 2016.
3. J. Domingo-Ferrer, L. Zhang and K. R. Choo, Cryptography in Cloud Computing: Practice,Theory and directions in future research," Future Generation Comp. Syst., vol. 62, pp. 51–53, 2016.
4. J.Weng, H. Lu,Y. Yang and Y. Yang, "Fine grained proxy re-encryption with sharing of data based on cloud," Pervasive and Mobile Computing, vol. 28, pp. 122–134, 2016.
5. D. Quick and K. R. Choo, "Google drive: Forensic analysis of data remnants," J. Network and Computer Applications, vol. 40, pp. 179–193, 2014.
6. B. Waters and A.Sahni, "Encryption based on fuzzy identity," Cryptology Advances- EUROCRYPT 2005, 24th Yearly International Conference on Cryptographic Techniques-Applications and Theory Aarhus, Denmark, May 22-26, 2005, Proceedings, ser. Lecture Notes in Computer Science, vol. 3494. Springer, 2005, pp. 457–473.
7. K. Li, R. H. Patterson and B. Zhu, "Disk bottleneck avoidance in data domain system," in 6th USENIX Conference Storage and File Technologies, FAST 2008, February 26- 29, 2008, San Jose, CA, USA. USENIX, 2008, pp. 269–282.
8. S. Keelveedhi, T. Ristenpart and M. Bellare, "Secure deduplication with Message-locked encryption," in Advances in Cryptology -2013 EUROCRYPT, 32nd yearly International Conference on the Applications and Theory of Cryptographic Techniques, Athens, Greece, May 26-30, 2013.
9. G.Segev, A. Raghunathan, and I. Mironov, "Encryption of locked messages for messages dependent on locks,"in Cryptology Advances - 2013 CRYPTO - 33rd Yearly Conference.
10. T. Ristenpart and S. Keelveedhi, "Encryption aided with server for deduplicated cloud storage," in Proceedings of the 22th USENIX Security Symposium, Washington, DC

AUTHORS PROFILE

Dr. R.Naresh [First author], Associate Professor in the Department of Computer Science and Engineering at SRM Institute of Science and Technology. He has many years of teaching and research experience. He has done B.E and M.E in Computer Science and Engineering from Anna University and done Ph.d in Computer Science and Engineering. His research interest includes Cloud Computing, Distributed Networks, Group Key Management, Web Security.

Vibhor Lohani [Second Author], Currently in final year Bachelor of Technology in Computer Science and Engineering at SRM Institute of Science and Technology. His main area of interest includes Artificial Intelligence, cloud computing and python programming. He has done internship in Infosys and has undergone training in ONGC and DRDO.

Naman Agarwal [Third Author], Currently pursuing Bachelor of Technology in Computer Science and Engineering at SRM Institute of Science and Technology. This paper is one of his beginning works in research area. His main area of interest includes Machine learning, image processing, internet of things and cloud computing.