Article Citation Format
Oloyede, A., Adedara, O., Nwaocha Vivian .O., Oduroye, A., Ogunlana, D., Adeyemi, B. & Longe, O. (2020). Development of a Secured Authentication Technique for Accessing De-duplicated Data from Private Cloud Using One Time Password. Journal of Advances in Mathematical & Computational Sc. Vol.8, No. 2. Pp 67-78

Article Progression Time Stamps

Article Type: Research Article
Manuscript Received 27th May, 2020
Final Acceptance: 19th June, 2020
Article DOI Prefix: dx.doi.org/10.22624

Development of a Secured Authentication Technique for Accessing De-duplicated Data from Private Cloud Using One Time Password

[1]Oloyede, A., [2]Adedara, O., [4]Nwaocha Vivian .O., [3]Oduroye, A., [4]Ogunlana, D., [5]Adeyemi, B. & [6]Longe, O.
[1,3,5]Department of Computer Science Caleb, University, Imota Lagos, Nigeria.
[2]Department of Computer Science, The Federal Polytechnic, Ado-Ekiti, Nigeria
[4]Department of Computer Science, National Open University of Nigeria, (NOUN) Abuja, Nigeria.
[6]School of IT & Computing, American University of Nigeria, Yola, Nigeria

ABSTRACT

The main aim is to de-duplicate the redundant files in the cloud and also to improve the security of files in public cloud service by assigning privileges to the documents when it is uploaded by confidential user. Methods: To achieve the objective the authors have used the AES algorithm to encrypt the file stored after de-duplication in the cloud. De-duplication is done based on comparison of contents, file type and size. For an authorized user to access the file from the cloud, generation of OTP using SSL protocol is adopted. Findings: Files uploaded in the cloud are encrypted using traditional encryption algorithms which don't provide high levels of security. Files can be accessed by anyone who is authorized. Privileges are not considered. During de-duplication, only the name and size of the files are considered. Application: Files within the public cloud can't be viewed by everyone who has registered with the cloud. Those who have the respective privileges can only view the file. Proof of Ownership is assured. Since de-duplication is done based on the content redundancy within the cloud storage is avoided. Usage of OTP ensures that the content is viewed by the individuals who have the respective privileges related to the file. These concepts provide additional security to the files stored in the public environment.

Keywords: AES, De-Duplication, Duplicate Copies, OTP, Privileges

1. INTRODUCTION

Cloud computing technology is used to store enormous amount of data and appear to be a virtual resources to the users. It is dynamic and can be easily accessed from anywhere provided with internet. It encapsulates the platform and execution details from the user. Instead of using costly hardware components, cloud service is comparatively cheap. It is extensible, scalable and updated with ease. Ex: If the user currently has 2GB of space and is in need of further storage space (Li, 2013; Itani, 2009), he can expand it easily. Private cloud provides more security (Mohan, 2013; Popović, 2010; Prakash, 2012) with less storage space. It can be accessed easily.

It is suitable to use within the organisation. Data (Annamalai, 2015) can be accessed based on privileges. The keys for the files stored in public cloud are usually stored in private cloud. On contrary, public cloud provides data storage with less security. To secure the data from losing its confidentiality, privileges (Prakash, 2015; Annamalai, 2015) are given to the files, so that only specific people can access the file. Privileges can be given both the types of cloud service (Saravanakumar, 2012). Authorization to the cloud is provided based on the credentials stored in the database during registration with the cloud. De-duplication is a data compression technique used to eliminate the redundant copies in the cloud enhancing (Mell, 2009; Khan, 2016) the storage capacity. In file level, it eliminates the duplicated files and at block level redundant blocks in the file are eliminated in non-identical files. The file attributes like size, content and type are checked. Privacy concerns are present due to insider and outsider attacks.

Data are encrypted (Corena, 2012; Ryan, 2011) for security (Pearson, 2013; Subashini, 2011) reasons. In traditional encryption, when the same file is uploaded by different people different cipher texts are created for each individual. This makes de-duplication difficult. This key is used to encrypt or decrypt the file. Keys are present with the user and the cipher texts thus generated are stored in the cloud. Here the cipher text produced for identical copies of file will be same and helps in de-duplication. Proof of ownership is provided to the files to ensure the user holds the file in spite of duplicate copies. When convergent encryption is used de-duplication of cipher text is possible and proof of ownership helps to enhance confidentiality. Issues arise when de-duplication with privileges are tried to be implemented at same time.

2.RELATED PREVIOUS WORKS

Venkatesh, Sharma, Desai et al. (2014), aimed to minimize the data duplication along with data. Security is to protect the confidentiality of the data. The security is provided by using many encryption techniques to encrypt the data before outsourcing. The users are checked whether they are authorized or not. Encryption is symmetric. SCSP is used to reduce de-duplication. The algorithm used here is novel encryption key generation algorithm. During the process of uploading the file, a tag is generated. It helps to identify the duplicates. These tags are stored in a separate table. The project has the advantage that the system is suitable for backup storage by using authorized de-duplication. In data duplication the encrypted keys are generated by private key cloud server. The proposed system is storing the authorised privileges.

Kumaresan and Visuwasam (2015) had tried to overcome the disadvantage of traditional encryption techniques using convergent encryption. Hash calculation is done at block level. If the target device finds a duplicate, then it doesn’t store a duplicate block. Instead it references to the existing block. Data Duplication is an important technology used in many companies to save a lot of money on storage cost and bandwidth by avoiding the replication. This paper proposed that the server will pop-up the duplication message if a file duplication is found. The security is provided by encryption algorithm. In order to protect the higher secured data, the effort made to increase the data accuracy check and hardware utilization. RSA algorithm is used to increase the security in hybrid cloud. Walunj, Lande, and Pansare (2014). Had said that the data duplication plays a vital role in eliminating the replicated data copies. This process eliminates the duplicate copy by saving only one copy and replacing the other copies with pointers. The pointers make the link to original copy. De-duplication widely used in companies for backup and disaster recovery process. The paper aimed to provide the authorized de-duplication check by combining the convergent encryption algorithm in hybrid cloud. Hence to access the data in public cloud the user must provide the key.

Li, Li, Chen et al. (2015) say that the same file may be saved in several different places by different users, or two or more files that aren't identical may still include much of the same data. The convergent encryption technique is used to encrypt the data. Proof of data is proposed by them. It is used while uploading file. If there is a copy of the uploaded file along with same privilege duplicate could be found by the user. The duplicate-check tokens of files will be generated by the private cloud server with private keys. Based on user's identity secret key will be provided. IBS (Identity Based Signature) is a method in which user will sign the message. It is used to avoid the issue of fragmentation of images.

3. THE DEVELOPED APPROACH

Cloud computing is ruling the progressing hi-tech environment. It helps to store voluminous data and access can be gained by users from anywhere. The service providers provide the storage and access to people at low costs. With more number of people started to use cloud, there can chances of redundancy. People belonging to same organisation can upload a file with same content with different names. This can occupy a lot of space. De-duplication helps to reduce the redundant copies of files in the cloud. When files with same content are found, they are de-duplicated. The data owner can assign privileges for the file so that only specific people can gain access to it. The file will not be available for others. This is one of the advancement related to security for the cloud service. Security is provided to the data in public cloud.

The data stored in cloud are de-duplicated based on the content, size and type using AES algorithm. This enhances the confidentiality of the files within the cloud. When an authenticated user tries to access a file, his details are checked with the database. If the credentials match with the database, then an OTP will be sent to the client's mail. The client will be requested to enter the received OTP on the web interface. The previously generated OTP is checked with the value the client entered. If both the values match then access to the file is given to the client. Client can access the file only if his privilege is available in the list of privileges given to the file. It provides increased security to the data within the public cloud.

3.1 A Framework for Accessing De-Duplicated Data

De-duplication helps to reduce the redundancy in the cloud. Instead of saving many copies of same file and increasing the storage space, it's better to have a single copy of file and pointing the reference to the file. Though most of the issues in cloud aren't solved yet, one of the major challenges (Wei, 2010) is providing security to the file in public cloud. When the data owner tries to upload the file, he could assign the list of privileges to the file. When an authenticated user tries to access a file, he could download it only when his privileges are available in list of privileges attached to the file when the file is uploaded. He will be sent an OTP which he should enter in the web interface. AES algorithm is used to encrypt the file within the cloud. The entire system provides a conduit to store voluminous data with confidentiality element based on privileges. As shown in Figure 1, a new user first registers with name, password, email and designation to utilize the cloud service. All the credentials are stored in MySQL database. The designations are included by an administrator. He assigns the appropriate privileges for the designations. The new user should select one amongst the designations included by the administrator.

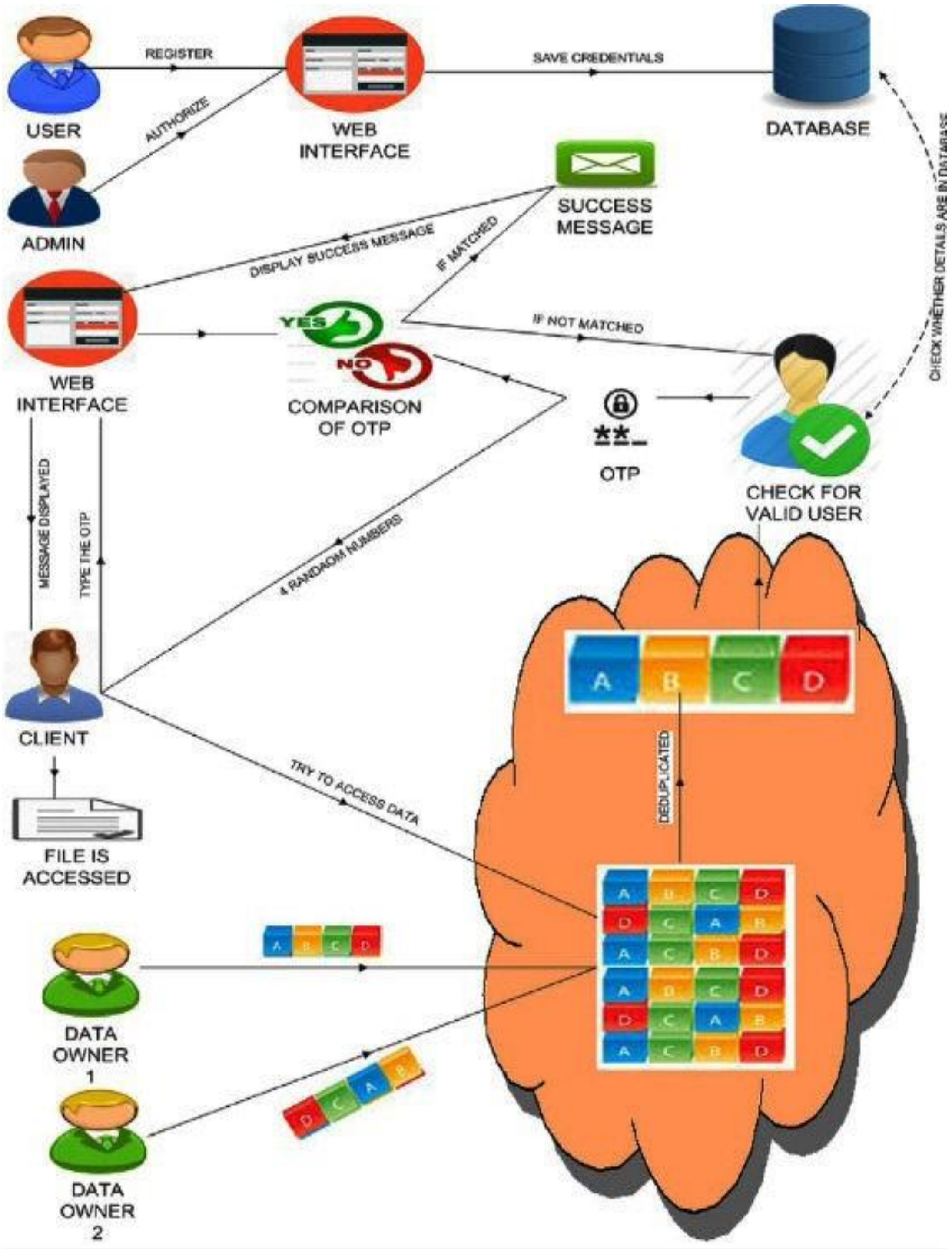

Figure 1: A Frame Work for Accessing De-Duplication Data

He is present to maintain the credentials stored in the database. Once the user registers, he can login to the cloud. The authorized users who upload the data are the data owners. These data owners can upload the file and provide privileges to the file. Based on the privileges the clients can access the file. Many data owners can upload the files with same content but different file names. Based on size, content and type the files are de-duplicated. The de- duplicated files are encrypted using AES algorithm and stored within the cloud. The client tries to access the file stored within the cloud. Validation of users is done based on the credentials stored within the database. If the credentials of client match to the details in the database, then an OTP is generated and sent to the client's email address. The client enters the received OTP into web interface.

The previously generated OTP which is stored in database is checked with the user entered OTP. If they are same then a success message pops up and the file can be accessed by the client. Else an error message is displayed.

3.2 Assigning Privileges to Designations

The administrator can login and include the necessary designations, the privileges. He can also assign the appropriate privileges to specific designations as in Equation 1. On the user's side, users should first register. Once they have registered, they can login with the required credentials. Based on the designation they give during registration, files will be made accessible to them:

$$Ad\min \leftarrow \mathrm{login} \cup \mathrm{Designation} \cup ACCESS \quad privilege \qquad (1)$$

3.3 De-Duplication and Encryption

When the data owner tries to upload a file, he should include the necessary privileges to the file. On uploading the file is checked for duplication based on the content, size and type. Finally, the file is encrypted by using AES algorithm and saved in the cloud as in Equation 2 and Equation 3:

$$FileUploaing \leftarrow file\,\mathrm{Pr}\,ivilege \cup fileCharacteristics \qquad (2)$$

$$FileCharacteristics \leftarrow AES(Contient \cup Size \cup Type \qquad (3)$$

3.4 OTP Generation

When an authenticated user tries to access a file within the cloud, the privileges of user are checked whether it matches with the requestor's privilege or not. If it matches an OTP is generated using SSL protocol and sent to the requestor's mail address which is given at the time of registration. The client should enter this value into the web interface. When both the values match, then the file can be accessed by the client. Else an error message is thrown to the client as in Equation 4

$$OTPGeneration = \left\{ \begin{matrix} True, filecharacteristic = \underset{otherwise}{Matched} \\ {}_{false} \end{matrix} \right\} \qquad (4)$$

3.5 Algorithm and Flow Process of Proposed De-Duplication

The algorithm for the de-duplication is discussed below. If the Cloud Service User request uploads the file then Checks for the Identity belongs to the same organization otherwise the map-reduce action is performed:

```
De-duplication()
begin
Cloud Service User (CSU) request the cloud service ($CSU_{req}$);
Cloud Service Provider (CSP) provides the appropriate Services ($CSP_{res}$);
if(CSUreq == "file upload")
identify the request source ($Id_{req}$) if($Id_{req}$
    == "same_organization")
    preprocess the request;
    else
    peform map-reduce;
    goto L1;
else
    access the $CSP_{res}$ from cloud
L1: Categorize the contents ($Cat_{content}$);
if($cat_{content}$ == "file content")
    identify the redundant block;
    eliminate the redundant bloc;
else
    identify the file name;
    identify the content of a file;
    perform map-reduce operation for the file content end
```

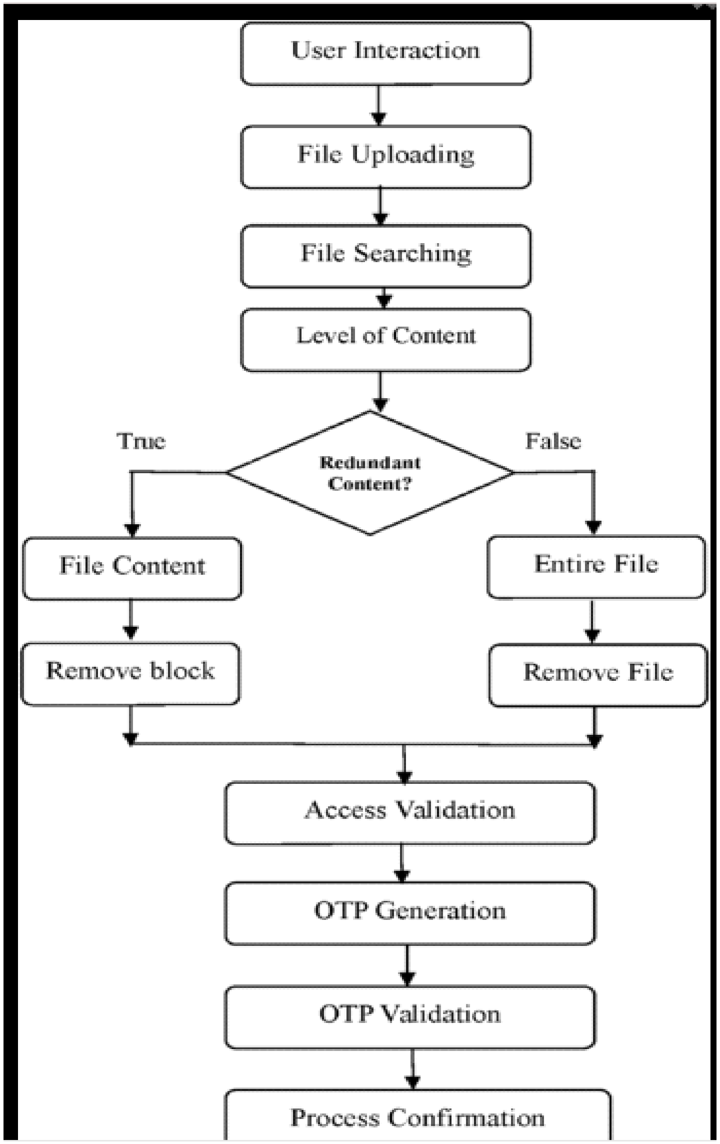

Figure 2: De-duplication process flow that how the files are uploaded and how the OTPs are generated

4. RESULTS AND DISCUSSION

The experimental setup of the proposed algorithm using Hive. It is a data warehouse tool which is used to handle structured data. This tool follows the Hadoop and Bigdata technology for analysing and summarizing the data with effective manner. This tool also uses the features like OLAP, HDFS support, HIVE QL with scalable and effective manner.

VM player (Liu, 2010; Bu, 2013; Tian, 2015) is used for creating VM's. The VM (Tanahashi, 2010) has been started with various setting by using Ubuntu 14.04 LTS. Initially the Hadoop is installed and create the cluster with one node which is act as Distributed File System. The Map Reduce program has been tested. The Hive act as a data warehouse for storing the data. The Hive directories are created in Hadoop for effective management of data. The Map-reduce programming has been implemented for identifying redundant data. The file redundancy is checked with file name along with file contents, whereas the file content with different file name is handled in different manner.

There are two stages to perform redundancy check in an entire file. In first stage, it checks the file name and in second stage it applies the Map-reduce for identifying a unique content. Compared to the existing system the encryption algorithm AES used in the proposed system proves to be much secured. In previous systems, privileges for accessing file are not included. In the proposed system, privileges for each file are included. New privileges can also be added and controlled by an administrator. Differential authorisation is included. The major advancement of the proposed system is the generation of One Time Password using SSL protocol for each user validation before accessing the file.

De-duplication takes place based on the content, size and type. Traditional SCSP database are not used in the proposed system. The File tag (TAG) computes the SHA-1 has of the File. The Encryption Algorithm AES with 256 bits, encrypts the different ranges of file from 10 – 500 MB of different category of files as shown in Figure 3. At the time of file initialization there won't be any duplication and the file transfer in the public cloud also disclosed. When the data is uploaded in the cloud, if there is any duplication of the data's is identified those duplicated data can be removed. Figure 3 shows if the file size is increased then for authentication the data encrypted is also gradually increased

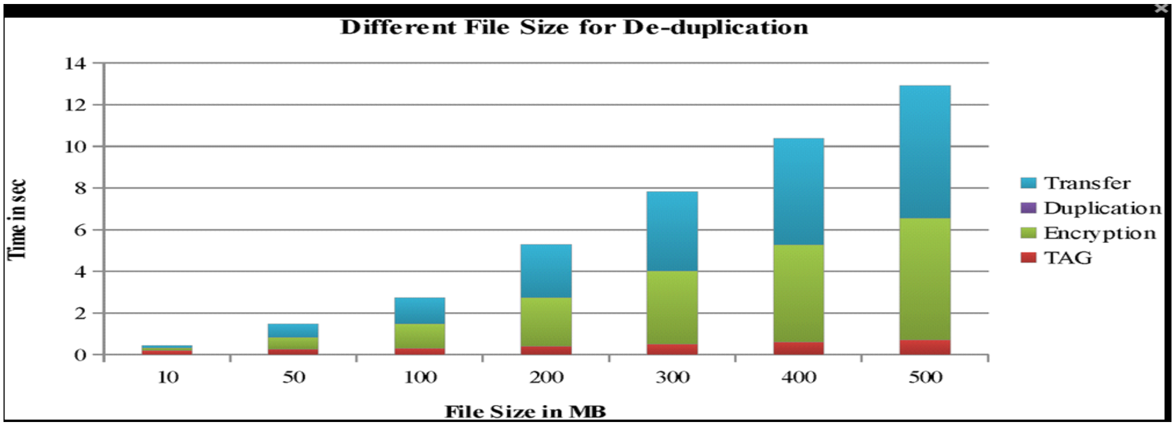

Figure 3: Breakdown for different file size

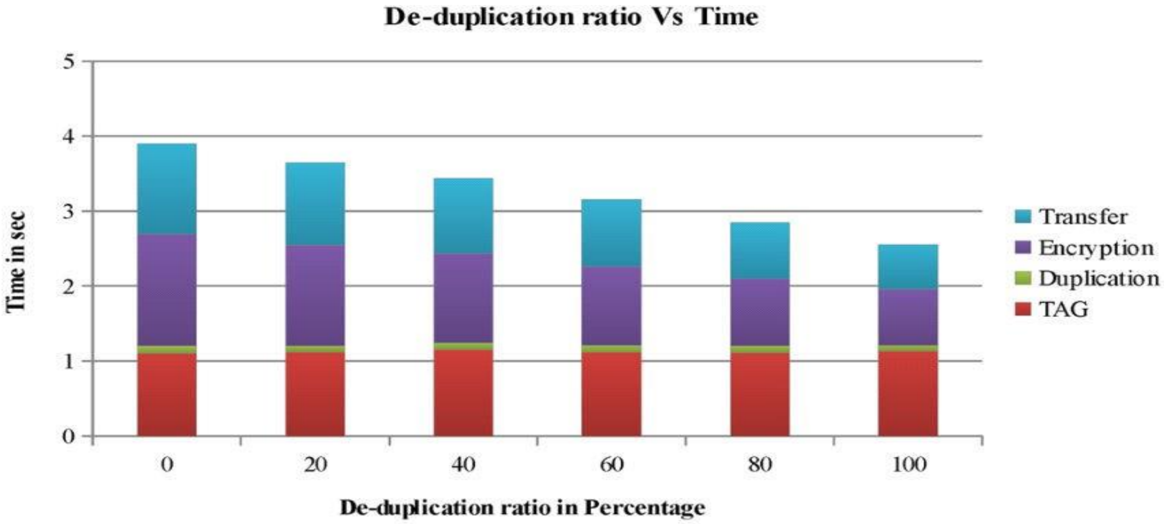

Figure 4: De-duplication Ratio with time

After completion of the de-duplication in Cloud storage a sample of 5 different files of different size has taken and compared its Similarity and Redundancy rate. It has absorbed clearly that for the very lesser in file size has more than 60% of Similarity Content. This shows that maximum number of user creates the file size with the lesser in file size and the File 5 as shown in Figure 5 shows that higher percentage of similarity in file contents whereas the redundancy data rate for all the file content remains very negotiable. This clearly indicates that de-duplicated content has been removed.

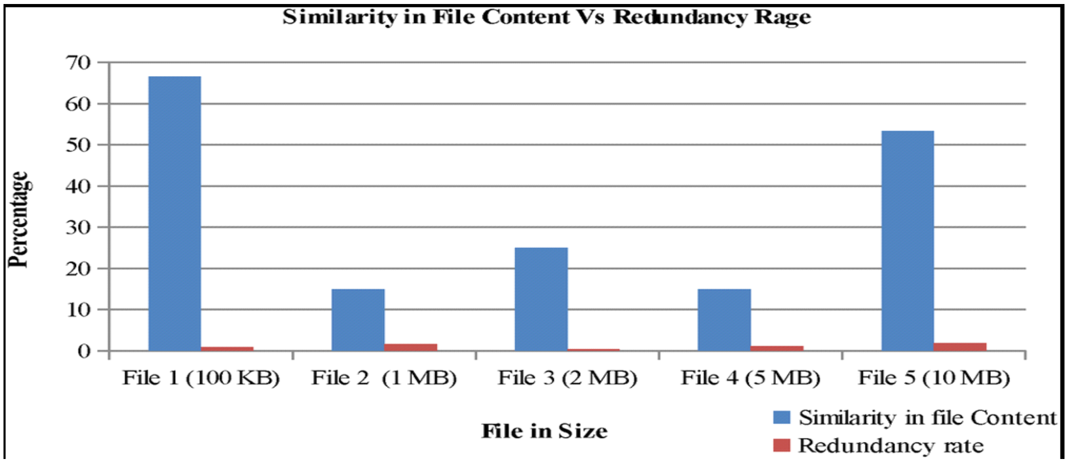

Figure 5. Comparative between similarity in file content and redundancy rate in percentage

5. CONCLUSION

Cloud computing had emerged as a boon to the advancing world. Though most of the issues have not been solved the major concern of providing security for the data within a public cloud is solved by the proposed system by issuing differential authorization to the files uploaded into the cloud. Further access of files by the clients are authorised by OTP method. Implementation of a prototype related experiments were conducted. It is shown a minimal overhead is incurred compared to the existing systems.

EndNote:
This paper had previously been published in the Journal of Digital Innovations and Contemporary Research in Science, Engineering & Technology. Copyright duly obtained and certified

REFERENCES

[1] Bu, X., Rao, J., & Xu, C. Z. (2013). Coordinated self-configuration of virtual machines and appliances using a model-free learning approach. IEEE Transactions on Parallel and Distributed Systems, 24(4), pp. 681–690. doi:10.1109/TPDS.2012.174

[2] Kumaresan, G., & Maria Michael Visuwasam, L. (2015). Enhanced In-Line Data Duplication and Secure Authorization in Hybrid Cloud. International Journal of Innovative Research in Science. Engineering and Technology, 4(2), pp. 2319–8753.

[3] Li, N., Zhang, L. J., Xu, P., Wang, L., Zheng, J., & Guo, Y. (2013, June). Research on pricing model of cloud storage. Proceedings of the 2013 IEEE Ninth World Congress on Services (pp. 412-419). IEEE. doi:10.1109/SERVICES.2013.70

[4] Liu, Q., Weng, C., Li, M., & Luo, Y. (2010). An In-VM measuring framework for increasing virtual machine security in clouds. IEEE Security and Privacy, 8(6), pp.56–62. doi:10.1109/MSP.2010.143

[5] Mell, P., & Grance, T. (2009). Effectively and securely using the cloud computing paradigm [PowerPoint]. NIST, Information Technology Laboratory.

[6] Mohan, P., & Thangavel, R. (2013). Resource Selection in Grid Environment Based on Trust Evaluation Using Feedback and Performance. American Journal of Applied Sciences, 10(8), pp. 924–930. doi:10.3844/ajassp.2013.924.930

[7] Pearson, S. (2013). Privacy, security and trust in cloud computing. In Privacy and Security for Cloud Computing (pp. 3–42). Springer London. doi:10.1007/978-1-4471-4189-1_1

[8] Popović, K., & Hocenski, Ž. (2010, May). Cloud computing security issues and challenges. Proceedings of the 33rd International Convention, Opatija (pp. 344-349). IEEE.

[9] Prakash, M., Farah Sayeed, R., Princey, S., & Priyanka, S. (2015). Deployment of Multicloud Environment with Avoidance of DDOS Attack and Secured Data Privacy. International Journal of Applied Engineering Research, 10(9), pp. 8121–8124.

[10] Prakash, M., & Ravichandran, T. (2012). An Efficient Resource Selection and Binding Model for Job Scheduling in Grid. European Journal of Scientific Research, 81(4), pp. 450–458.

[11] Ryan, M.D. (2011). Cloud computing privacy concerns on our doorstep. Communications of the ACM, 54(1), pp. 36-38.

[12] Samadhu, A.A., Rambabu, J., Pradeep Kumar, R., & Santhya, R. (2015). Detailed Investigation on a Hybrid Cloud Approach for Secure Authorized Deduplication. International Journal for Research in Applied Science & Engineering Technology, 3(2), pp. 226–269

[13] Saravanakumar, C., & Arun, C. (2012). Traffic analysis and shaping of the cloud services over common deployment model using cloud analyst. International Journal of Computers and Applications, 43(4), 33–37. doi:10.5120/ pp. 6094-8281

[14] Tian, W., Zhao, Y., Xu, M., Zhong, Y., & Sun, X. (2015). A toolkit for modeling and simulation of real-time virtual machine allocation in a cloud data center. IEEE Transactions on Automation Science and Engineering, 12(1), pp. 153–161. doi:10.1109/TASE.2013.2266338

[15] Venkatesh, B., Sharma, A., Desai, G., & Jadhav, D. (2014). Secure Authorised Deduplication by Using Hybrid Cloud Approach. International Journal of Innovative Research in Advanced Engineering, 1(10), pp. 221–227.

[16] Walunj, R. S., Lande, D. A., & Pansare, N. S. (2014). Secured Authorized Deduplication Based Hybrid Cloud. International Journal of Engineering Science, 3(11), pp. 34–39.

[17] Wei, Y., & Blake, M. B. (2010). Service-oriented computing and cloud computing: Challenges and opportunities. IEEE Internet Computing, 14(6), pp. 72–75. doi:10.1109/MIC.2010.147