I. J. Computer Network and Information Security, **2024, 1, 48-60**
Published Online on February 8, 2024 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijcnis.2024.01.04

Secure Access of Folders and Files after Removal of Duplicacy over the Cloud

Deepika Gautam*
Department of Computer Science, Babasaheb Bhimrao Ambedkar University, Lucknow, 226025, India
E-mail: deepika.gautam.lko@gmail.com
ORCID iD: https://orcid.org/0000-0002-1056-3160
*Corresponding author

Suvendir Rimer
Department of Electrical and Electronic Engineering Science, University of Johannesburg, South Africa
E-mail: suvendir@uj.ac.za
ORCID iD: https://orcid.org/0000-0002-4976-9721

Vipin Saxena
Department of Computer Science, Babasaheb Bhimrao Ambedkar University, Lucknow, 226025, India
E-mail: vsax1@rediffmail.com
ORCID iD: https://orcid.org/0000-0003-1035-1704

Received: 17 April 2022; Revised: 30 May 2022; Accepted: 18 July 2022; Published: 08 February 2024

Abstract: Cloud Computing has been the most popular approach of computing due to faster access to folders and files at a low cost. Hence, many organizations are shifting the old long database folders and files over the cloud which may be text, audio, video or in the other formats. Due to large size of the database with multiple storages of folders and files over the cloud, there may be chances of duplicate access of the database folders and files which may cause the loss of time of execution or accessing the database files. In the present work, a technique is developed to remove duplicate files in the form of .txt, .doc, .jpg, .pdf as well as duplicate folders after applying a well-known ElGamal algorithm later on converted as fuzzy ElGamal technique, for faster retrieval of files in a very secure manner. For this purpose, Unified Modelling Language (UML) model is developed which has been implemented through Python programming language. The computed results towards the model's efficiency have been depicted through tables and graphs, on a large database in the form of folders and files of Indian railway reservation system. The present work is significant for the large organizations and also useful for the users working over the cloud for faster accessing of the folders and files.

Index Terms: Cloud Data Security, Encryption, Decryption, ElGamal Algorithm, Fuzzy Logic, Cryptography.

1. Introduction

A pre-plan plays a vital role in developing any system model, whether it is related to the hardware or software models. The model gives information on how every entity communicates to each other and the blueprint is called as a system model. It enhances the visual representation of the system; otherwise, a wrong model leads to failure of the system, hence it is necessary to construct the model efficiently. The system model helps in authentication, integrity, interdependencies, and steadiness of the system to design a complex research problem. In this work, a system model is developed for accessing cloud servers for the complex database files of the Indian railway reservation system and depicted in the Fig.1.

For efficient storage of complex database files, a torus topology is used to arrange servers over the cloud and is called as cloud storage. The role of torus topology is to connect finite numbers of servers wrapped around the mesh and if k dimensional servers are wrapped around mesh, then it has 2k neighbor's servers. If the folders and files are not available on the assigned server, then it will search from the neighbor servers. This topology is very cost effective and decreases the parallel communication and reduces the timing of search of desired information. One can access the desired file in a minimum time, as depicted in the Fig.1. The complex database is selected from the Indian railway reservation department. The authorized administrator is loading a huge amount of database in the form of folders and files over the cloud servers through a high-speed Internet gateway. The administrator does not provide the authentication, but it is

proposed by the third party whose responsibility is to provide a private key to the authorized user who has public key information for accessing the unique folders and files from cloud storage. Third party authentication is always done for the users who have account over the Google, Facebook, Twitter or any other social networking websites. In this technique three media accounts work together for authenticating the authorized users. In the figure, users are considered as working in the various organizations, individual user's and other categories of the users who are accessing the cloud storage across the cloud around the globe. In this work, the reasons for selection of cloud storage are that the stored folders and files may be used for the cloud computing.

Cloud's massive data require a reliable and robust protection shield hence cryptography plays a vital role for security of the data files. In the literature, various kind of cryptography algorithms are proposed by the researchers in which encryption algorithm makes plain text into a non-readable form known as cipher text. It is complex and hard to convert to plain text and for this purpose further decryption algorithm, uses a key to access cipher text via private key. In the present work, for the security of the stored database, the ElGamal and fuzzy ElGamal techniques are used for the generation of private and public keys and can be stored between the authorized user and the third party who is providing the authentication services. For reduction of accessing time, torus topology is used for the arrangement of servers inside the cloud that can be used as cloud data storage.

The Indian railway system has been functioning in India since 2002. Initially, the tickets were in simple text format using C and Fortran as a front-end high-level programming language with back-end support of structural database system. Due to the Graphical User Interface (GUI) evolution, object-oriented software was developed and is now proprietary of Center for railways information Services (CRIS). It supports text, audio, video files. Since the Indian railway reservation system is much vast, the database contains very large complex files related to finite number of users. The above system model is proposed for the faster accessing of complex database files in unique form and secure manner.

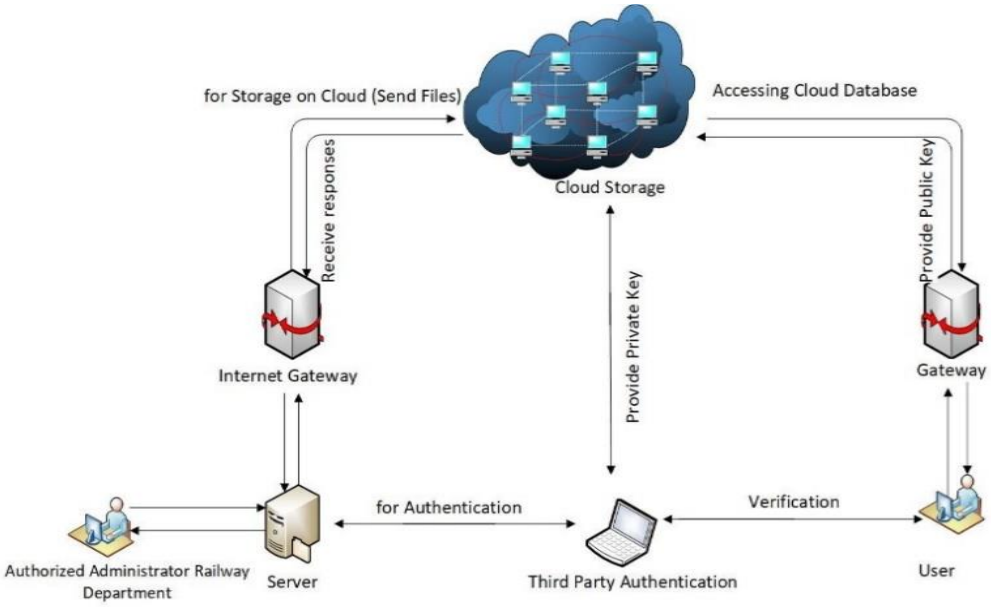

Fig.1. System model for accessing the complex data files

2. Review of Literature

The concept of security of information is ancient and day by day, it is going more popular due to the online transactions through handheld devices. In the early days, the file size was very small i.e., simple text. Users can easily transfer text information with limited size from one place to another in synchronous and asynchronous mode. But due to rapid changes in technology, the file size of data increasing exponentially, and the type of information may be text, audio, video files etc. The bulky size of information to be transmitted from one device to another either through one to one, many to one, one to many and many to many communication techniques, there is a need of complete security system. The transmitted information must be secured for which different authentication and authorization techniques are available in the literature. Some have been cracked down by the intruders while others have excellent security features used in the banking sector. Due to the growth of file size from time to time in exponential form, there is a need for cloud storage and one can't keep the large files on handheld devices which may be a laptop, mobile, tablets, personal digital assistants, smart watches etc. Due to this aspect, cloud computing is the most promising area of research. It needs proper security so that the information of an organization must be secured and be transacted from device to device through the cloud in a secure manner. Many researchers have proposed new techniques for security of information treated as a cloud database, hence there is need to describe systematic development of research during the past years on cloud database security.

In the year 2016, the authors developed a novel approach through linear programming method and generated a

heuristic algorithm to minimize the threat levels for cloud data security by (Kang et al. [1]). Architecture for portable data possession (PDP) was derived from storing the data initially on an untrusted server and further developed a model for remote data checking and converted the untrusted server into the trusted server through public-key cryptography by (Bhukya et al. [2]). For the Homomorphic Hybridization of Encryption technique (HHET), the authors proposed a combination of ElGamal and Paillier Encryption approaches for secured cloud data with private and public keys. The HHET approach was claimed as stretchy, cost-effective and secured delivery at the receiver side (Rao et al. [3]). It was observed from the literature that many Cloud Service Providers (CSP's) handle the data over the cloud. The CSPs steal the data and sell them to earn profit. Keeping an eye on this act authors used integrity verification methodology with encryption mechanism for deployment of data using ElGamal and SHA-256 algorithm by (Panimalar and Subhashri [4]). To protect sensitive data related to finance and other agencies from cloud operators' authors introduced security efficiency data distributions algorithms and efficient data algorithms to provide security and excellent performance by (Gai et al. [5]). The cloud data security infrastructures related to E-learning to M-learning framework are proposed by (Adejo et al. [6]).

In the year 2017, Cloud data can be accessed from anywhere, and for security services, authors developed authentication scheme and worked on several kinds of attacks (Soni et al. [7]). Cloud data was required to protect from cloud service providers. A novel verifiable auditing scheme was used to remove third-party auditor during outsourcing data by (Xiang et al. [8]). Several sensitive data were stored over the cloud, needed high data confidentiality. Cryptography algorithm DNA and ElGamal were used to transfer data between the owner and data user by (Thangavel and Varalakshmi [9]). ElGamal cryptography hyper elliptic curve, and gravitational search were used for integer selection, encryption and decryption, key selection by (Thiyagarajan and Rao [10]). Models of cloud i.e., deployment and service delivery are used by many businesses and organization that prefer more data security by (Kumar et al. [11]). Once the data was deployed over the cloud, the owner lost his control; hence the authors developed a model which was designed with a combination of encryption algorithm that distributed data over the cloud according to data sensitivity (Alsirhani et al. [12]). Several algorithms were used for cloud data security such as the ARIA Cipher algorithm with ElGamal algorithm (Kaur and Wadhwa [13]). The other related references in the year 2017 and 2018 are (Kaur and Devgan [14] and Soni and Mishra [15]), respectively.

In the year 2018, blending the two algorithms provide a new hybrid algorithm that gave more efficient security i.e., Two Fish algorithm with ElGamal for Encryption and Decryption between two parties (Vedaraj and Prem [16]). Cloud handles a massive amount of data; therefore, authors discussed the necessity of a different storage and retrieval of big data with high security (Reddy [17]). Clouds had a host to store data. It can change storing data on different nodes and has fewer chances of data leaking (Londhe et al. [18]). Security over the cloud was a very complex issue and one of the best ways to encryption data because it had a key and access permission, which was handled by IT Security Specialists (ITSS), resulting in high security and privacy (Hyseni et al. [19]).

In the year 2019, Database-as-a-service was a cloud service for storing and retrieving data. The paper presented a different way of data encryption without sharing keys on a database (Gahia and Alaoui [20]). Cryptography provided security over the cloud. RSA, ElGamal and Elliptic curve algorithms were compared based on key size and running time (Mallouli et al. [21]). Identity Access Management (IAM) attacked over cloud and trespassers accessed accounts. A new system combination of RSA with ElGamal and Paillier has implemented. The promising task in cloud computing was security. A method to protect data using Blockchain that helps users to trace malicious data access was presented (Pradeep et al. [22]). Combining the encryption algorithm with vertical fragment distributed data over clouds, which makes them meaningless. Java application for simulation helped to evaluate encryption and hybrid fragmentation. Comparing the existing solution with the evaluated solution provides a secure mechanism of data security (Moorthy and Baranidharan [23]). Using ElGamal and Hyper Elliptical Curve Cryptography (HECC) provided many key sharing techniques between two parties. HECC was used for key generation by point addition and point doubling (Awad et al. [24]). Blockchain was another technique used over cloud data security. It controls the flow of authorized access and modification of the server (Devi and Ganesan [25]). An efficient security architecture for mobile cloud data is proposed by (Sekaran et al. [26]).

In the year 2020, Cloud approached deduplication the algorithms for integrity and security like Convergent Encryption, the proof of ownership (Prajapati and Shah [27]). Public Key cryptosystem used for security and that's proved. Using the Public Key cryptosystem, authors used RSA and discrete log together (Tripathi et al. [28]). Due to high demand for the cloud because of its lower cost, high performance, availability and storage, security threats and security issues. The authors proposed reorganization of AES whose plain text was 128 bits and the critical size for encryption /decryption is 128 bits (Bindu and Reddy [29]). The CP-ABE model based on the quantum model provided security of a key, encryption/decryption. In this model, medical data which was large and raw was used. Results gave accuracy and efficiency (Singamaneni and Pasala [30]). Password authentication was widely used on the client-side author worked on password security and hashed using a hash function and converted it into a negative password. ElGamal used public and private key to encrypt and decrypt the messages. Encrypt negative passwords worked with a cryptography algorithm and shielded from attackers (Srivalli and Raghavulu [31]). IoT used the cloud widely to transfer data. The author proposed an improved ElGamal Cryptography which was difficult to crack (Mohan et al. [32]). An efficient distributed cloud-based storage (STaaS) provided extra storage space with security. Data was divided into parts and spread over a network. That data was difficult to access by outsiders (Moorthy et al. [33]). A new lightweight cryptographic algorithm was analysed mainly focused on evaluation time, the public key and private key, working efficiently to provide sound security and performance by (Shuvo et. [34]). A review paper on the data security for mobile

Cloud Computing was written by (Qayyum and Ejaz [35]).

In the year 2021, Encryption was one of the best methods used in cloud security. Two layers of encryption break the limitation give more security (Thabit et al. [36]). A security technique has provided the time stamp helped in key, Encryption and Decryption ADSS Protocol give security for attacks on ciphertext (Pachala et al. [37]). This paper was on multi-cloud hosting environment privacy and security. A hybrid approach consists of three modules as Byzantine Protocol, Depsky architecture and Shamir. This hybrid approach was compared to the SAML protocol and Hybrid gives the best results than SAML Protocol (Zhang et al. [38]). Cloud Property graph was an interface that showed how data flow in cloud apps and how much time it took to run. Different available technologies can be used it by AWS, AZURE, and Kubernetes (Thabit et al. [39]). Pandemic realized the importance of cloud data which had patients, medical data. Cloud helps in the treatment of Melanoma skin cancer. The authors presented Z-Score method, which allows for color correction. The ordered Binary number system and image permutation are used to maintain privacy permutation. A challenge-response game model was used for analyzing security (Rajput et al. [40]).

In the present work, four categories of files format like .txt, .doc, .jpg, .pdf are considered and group of these files may form the folders. The reason for selection of four categories of the files is that these files are usually used by the authorized users on the high-speed internet services available over the cloud storage in which finite numbers of the servers are arranged by means of arrangement of servers through torus topology. For examples, Google drives, drop box, Tresorit etc., are available on the cloud severs in which authorized users stored the bulky size of the files with variations of file size. Drop box can store maximum length of file size of 2GB, Google drive can store maximum length of 15GB and similar finite space is available on the cloud servers which may contain duplicate folders and files in the various storage drives with identical file names and even contents may be identical. Therefore, the proposed work is implemented for the removal of duplicate file names as well as similar contents of the two different files by the use of python programming language. Further a well-known, ElGamal method is used to secure access of an individual file via a high-speed internet facility from one place to another. In this work, significant results have been obtained through simulation by increasing the file size and transmission time is also computed and represented in tables and graph and later on fuzzy ElGamal technique is also applied, to achieve high performance rate.

3. Research Methodology

Data security is one of the prominent areas of research due to tremendous growth in the file size and communication from one end to another end on the high-speed network. The role of a communication link which may be strong or weak is another aspect for making a secure link. When a file is transferred from one device to another device using a communication link, it must not be duplicate and it must be unique at the time of transmission. Further, the topological structure also plays a crucial role for faster accessing of any database. Due to the increase in the file size and limited storage available on the server, it is a must to store many databases on the cloud servers. It is also observed that day by day, a massive amount of database is available on the cloud servers, which contain duplicate file names, may be stored in different folders, and the contents of two files may be treated as copied contents. Therefore, a methodology has been proposed for the removal of duplicate files and same content at the storage size i.e., at the size of the cloud server. The proposed algorithm for large file size is considered the names of files are in the form of .txt, .doc, .pdf, .jpg and same algorithm may also applicable for other formats of the files and algorithm nearer to Python programming language is given below:

Pseudocode 1: To remove duplicate files

```
begin
path = ask directory(title='Select A Folder')
for base, direcs, filess in os.walk(path):
filename = Path(os.path.join(base, file))
if filename.is_file():
fileread = open(filename, 'rb')
fileH=hashlib.md5(fileread)
if fileH not in unique:
unique[fileH] = filename
else:
os.remove(filename)
print('Duplicate files removed')
end
```

The above algorithm has been implemented in the python programming language in which first authorized users have to select a folder thereafter all roots, directories and files are visited and check the duplicate name of the file and duplicate contents inside the file. Once the duplicate file is found then it will remove automatically by the above algorithm and the latest update file will be stored as it is. Further, the above algorithm has also been represented through the Unified

Modeling Language (UML) activity diagram which is in the following Fig. 2.

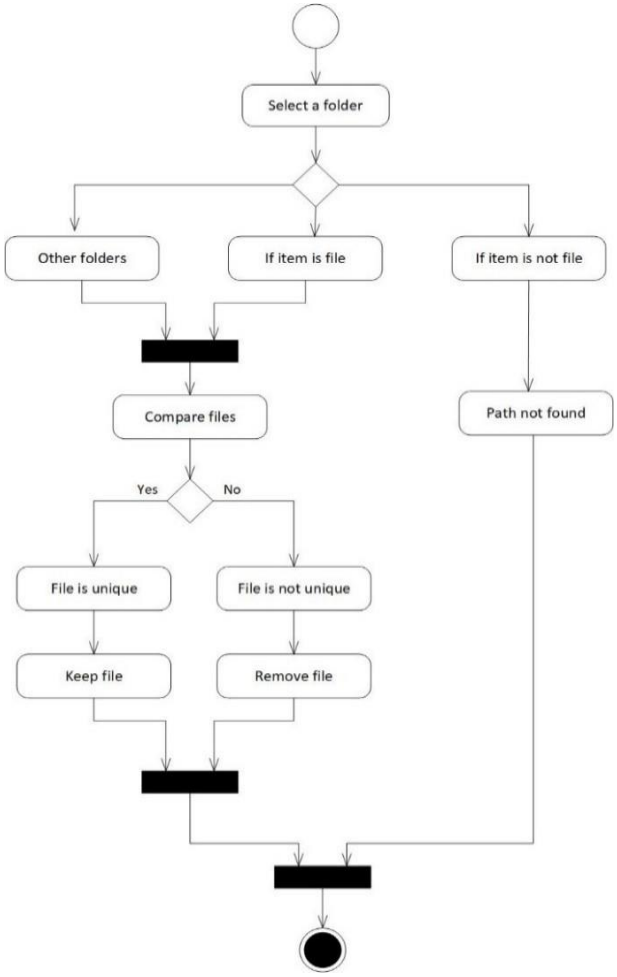

Fig.2. UML activity diagram for removal of duplicate files

From the above activity diagram, it is observed that folder is be selected first which contains many files formats as described above and if the desired files are available in the folder, then it will compare with the duplicates files (if available) and if duplicate files are not available in that folder, then it will search into the other folders till the duplicate files are found. The same interpretation is applicable for the duplicate folders also. When the individual files from the cloud server are to be transmitted securely, there is a need to apply an encryption and decryption algorithms. These algorithms may be categorized as symmetric or asymmetric key algorithm. A single key is treated as private or public in symmetric encryption and shared between sender and receiver. On the other hand, different private and public keys are used for asymmetric encryption.

The literature shows that asymmetric encryption and decryption algorithms are the more powerful techniques of security as compared to symmetric encryption and decryption. Therefore, the well-known ElGamal technique is used and pseudo code nearer to Python programming language is given below:

Pseudocode 2: To provide secure encryption and decryption via ElGamal Technique

```
begin
path = filedialog.askopenfilename()
file_extension = pathlib.Path(path).suffix
#Check file extension and read content
if file_extension=='.txt||.pdf||.doc||.jpg
content = input_file.read()
else:
print("wrong file");
init=time()
key generation(r)
key1=random.randint(pow(10,20),r)
encryption(msgs,r,i,j):
```

```
ke=gen_key(r)
s1=power(i,ke,r)
p1=power(j,ke,r)
decryption(ct,p1,key1,r):
i=power(p1,key1,r)
msgs=content
key1= key generation(r)
i=power(j,key1,r)
ct,p=encryption(msgs,r,i,j)
pt=decryption(ct,p1,key1,r)
print('\n Execution Time:',(time()-init))
end
```

The above pseudo code is compiled in the Python programming language. After removing the duplicate contents, the user selects a file whose file extension is first checked, whether the file belongs to .txt, .pdf, .doc or .jpg. According to the file extension, contents are read out from the file and converted to ASCII format before transmission over the internet. For secure transmission of the files, asymmetric ElGamal technique is used for generation of public and private keys. After the key generation, encryption and decryption are performed for secure transmission of the files. This is explained through following example:

Example of ElGamal Technique

Key Generation

1) Let prime number (pr) = 11, primitive root of pr is (j) = 2, random integer (z) = 8;
2) Compute A_1:

$$A_{1=}\, j^z mod\ pr\ =\ 2^8 mod\ 11\ =\ 3;$$

3) Then Get Public_key of the form (A_1, j, pr) = (3,2,11); and Private_key of the form (z,j,pr) = (8,2,11).

Encryption

4) Let message m = 5, Key(ke) = 9;
5) Using Public_key and Key(ke), Compute cipher text (C_1, C_2):

$$C_{1=} j^{ke} mod\ pr\ =\ 2^9 mod\ 11 =\ 6$$

$$C_2 = m.A_1^{ke} mod\ pr = 5 * 3^9 mod\ 11 = 9$$

Decryption

6) Proceeding with decryption on cipher text (C_1, C_2), then compute

$$x = C_1^z\ mod\ pr = 6^8 mod\ 11 = 4;$$

7) Compute the plain text

$$M = x^{-1} C_2\ mod\ pr = \left(\frac{1}{4}\right) * 9 mod 11 = 5.$$

where m=M =message.

Further the aforesaid ElGamal algorithm is converted by the use of fuzzy logic in which X is defined as non-empty set and a fuzzy set P on X having the membership function as $\mu \rightarrow P{:}X \rightarrow [0,1]$, where on P[X] is defined as degree of membership of element x which belongs to X ($x \in X$).

In the fuzzy logic, membership function maps all the elements of fuzzy set into the actual value lying between 0 and 1 further fuzzification process is done by transferring crisp value into linguist value and later defuzzification is done by converting into crisp value either 0 or 1. The pseudo code of fuzzy ElGamal technique is given below:

Pseudocode 3: To provide secure and enhance encryption and decryption via fuzzy ElGamal Technique

```
begin
key generation(r)
key1=random.randint(pow(10,20),r)
```

```
encryption(msgs,r,i,j):
ke=gen_key(r)
s1=power(i,ke,r)
p1=power(j,ke,r)
x=np.arrange(len(p1))
mfx=fuzz.trimf(x,(0,10,len(p1)))
decryption(ct,mfx,key1,r):
i=power(mfx,key1,r)
msg=content
key=gen_key(r)
ct,p1=encryption(msgs,r,i,j)
pt=decryption(ct,mfx,key1,r)
print('\n Execution Time:',(time()-init))
end
```

Further, the above pseudo code is compiled by the use of Python programming language for enhancing the security level via fuzzy based ElGamal technique. This technique is further described below:

Fuzzy ElGamal Technique

Key Generation

1) Choose any large prime number 'pr';
2) Select any integer as secret key(z), to be primitive root of mod pr (j);
3) Compute $A_1 = j^z \bmod pr$;
4) Then Get Public_key=(A_1,j,pr);

Private_key=(z,j,pr);

5) Convert Private_ key into fuzzy set and a fuzzy key will obtain.

Encryption

6) Choose a unique random number key(ke) between 1 to (p-1);
7) Using Public_key and key, compute cipher text (C_1, C_2).

Decryption

8) Firstly, defuzzify the fuzzy key into a Private_key;
9) Decryption proceeds on cipher text and plain text will obtain.

4. Results and Discussion

The above concepts for removal of duplicate files as well as duplicate contents have been implemented by taking the following parameters represent in table 1:

Table 1. Parameter used for simulation

Parameter	Value
Cloud Type	Public and Private
Total Users	50
File Size (KB/MB)	50,100,150,200,250 in KB,5,10,15,20,25 in MB
RAM	4 GB
Number of Processors	2
Processor Type	Intel(R) Core (TM) i3,2.40Ghz
Data blocks	500

The removal of duplicate files has been implemented on the 50 authorized users by simulating the data in the Python programming language with version 3.7 (64-bit). The type of cloud is considered both as public and private. The public clouds are opened to everyone while the private clouds are opened only to the authorized users. The file size of duplicate file is taken from 50KB to 25MB and for implementation purpose, users can select any size for removal of duplicate contents or duplicate files. The smallest unit of database is considered as 500 data blocks. A folder containing many duplicates files has been created using python programming language, as shown below in the following Fig. 3.

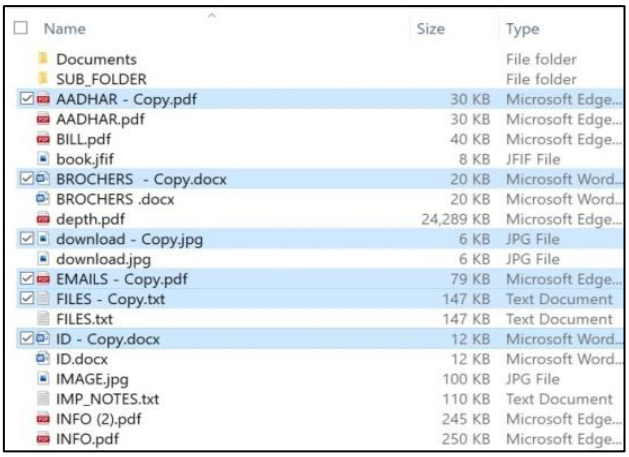

Fig.3. Folder containing duplicate files

After execution of the python code for removal of duplicate files, a new window is generated as depicted in the following Fig. 4. This folder contains all the four types of files which are considered in the recent work as .txt, .doc, .jpg and .pdf.

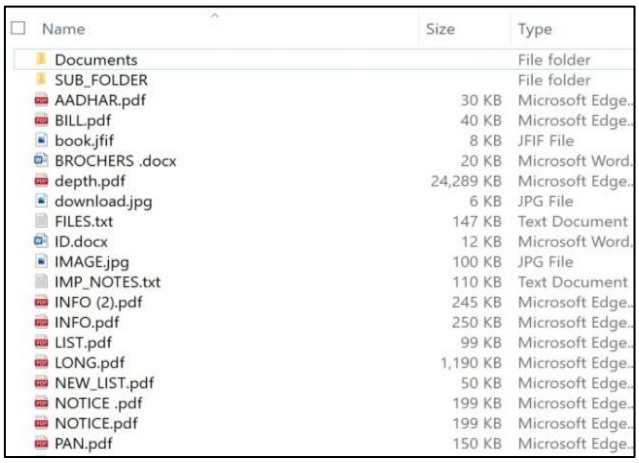

Fig.4. Folder after removal of duplicate files

Now, the above unique files are to be transmitted at the receiver end in a secure manner, thereafter ElGamal code is executed according to above-mentioned pseudo code and represented in the following Fig. 5 and 6 for 50 files of size 50kb and 250kb, respectively while Fig. 7 and 8 represent the implementation of fuzzy ElGamal technique for file size of 50kb and 250kb, respectively.

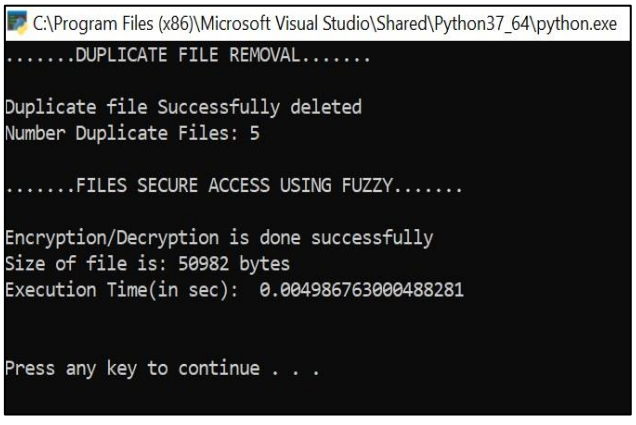

Fig.5. Results ElGamal code (File Size 50 KB)

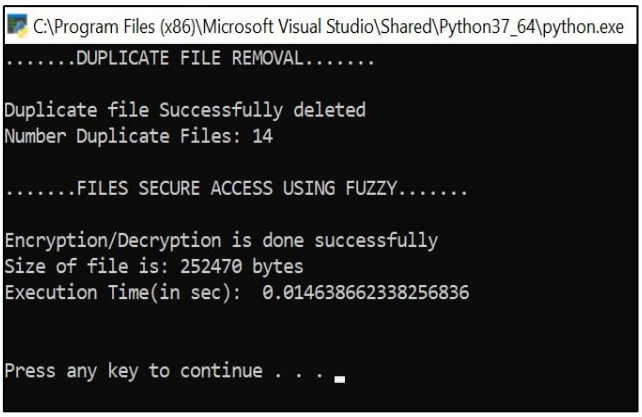

Fig.6. Results by ElGamal Code (File Size 250 KB)

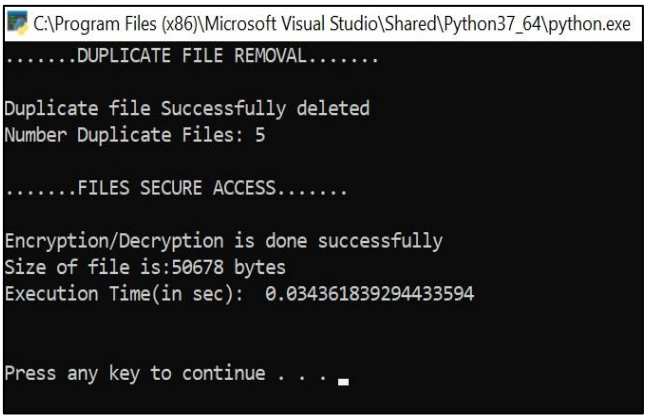

Fig.7. Results by ElGamal using Fuzzy Logic (File Size 50 KB)

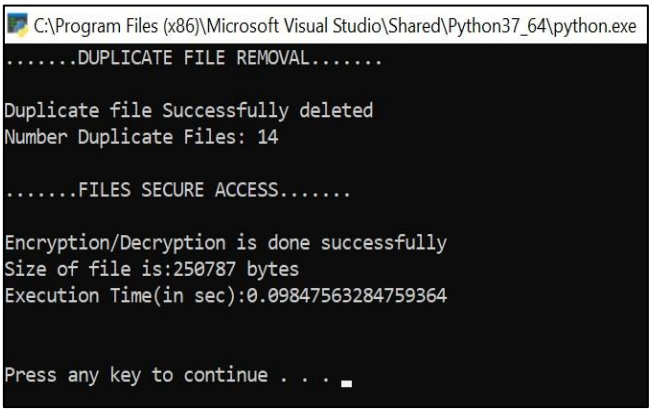

Fig.8. Results by ElGamal using fuzzy logic (File Size 250 KB)

In the above, two techniques i.e., ElGamal and ElGamal by the use of Fuzzy Logic have been compared in terms of response time as depicted in the following Table 2.

Table 2. Estimation of time through ElGamal and fuzzy ElGamal algorithms

Number of Duplicate Files	Encryption/Decryption of Selected File Size (in KB)	Encryption /Decryption Time (in sec)	
		ElGamal	Fuzzy ElGamal
5	50	0.03	0.004
7	100	0.05	0.006
9	150	0.06	0.007
11	200	0.08	0.009
14	250	0.09	0.014

From the Table 2., it is observed that when the file size is increasing then the estimated time is also increasing. The same response time is also represented in the following Fig. 9

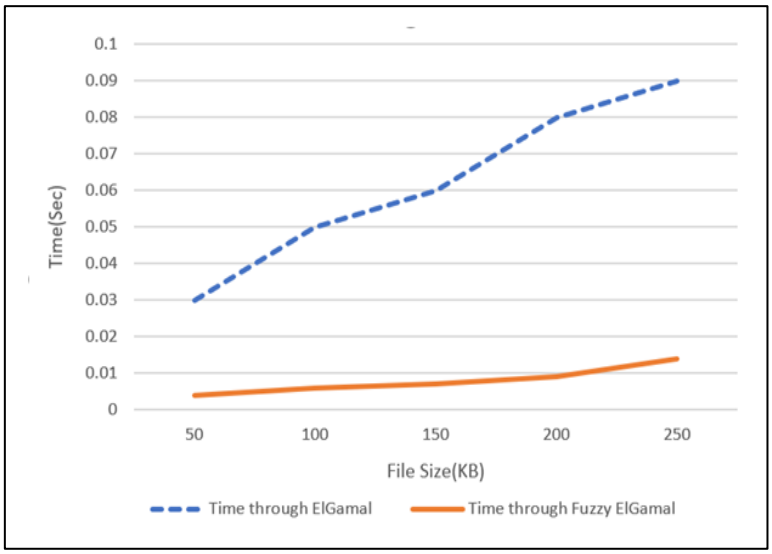

Fig.9. Response time through ElGamal and fuzzy ElGamal algorithms

The above graph is showing the two increasing lines. These lines depict that by increasing the file size response time also increases. The graph also compares the ElGamal and ElGamal with fuzzy logic algorithms. Time taken by ElGamal with fuzzy logics is much lesser than only ElGamal algorithm.

Such algorithms are necessary for the data security, data consistency and information assurance nowadays. ElGamal algorithm itself is a robust technique for protection, but using fuzzy logic in ElGamal algorithm encryption and decryption have become stronger and more reliable.

Table 3. Estimation of time for large file size

Type of file	Number of Duplicate Files to be Removed	Encryption/Decryption Selected File Size (in MB)	Encryption /Decryption Time (in sec)	
			ElGamal	Fuzzy ElGamal
.jpg	5	5	0.021	0.015
	7	10	0.078	0.045
	9	15	0.135	0.061
	11	20	0.184	0.093
	14	25	0.248	0.014
.pdf	5	5	0.118	0.038
	7	10	0.134	0.044
	9	15	0.145	0.079
	11	20	0.173	0.083
	14	25	0.189	0.092
.doc	5	5	0.144	0.068
	7	10	0.176	0.095
	9	15	0.201	0.122
	11	20	0.217	0.148
	14	25	0.234	0.175
.txt	5	5	3.153	1.749
	7	10	5.235	2.727
	9	15	7.588	3.615
	11	20	8.315	5.274
	14	25	9.563	6.796

Further a file for encryption and decryption can be of any type. Different file types take different response time. Table 3. is indicating .jpg, .pdf, .doc and .txt file of large file size and respective response times. It is observed from the table that in any type of file as far as file size increases, response time of secure file access is also in increasing manner. The above table is created for four kinds of file size but when the size of file will increase the response time will also increase in both the algorithm but fuzzy based ElGamal reduces the response time in comparison of normal ElGamal Technique.

5. Conclusions

From the above, it is observed that cloud server contains huge number of files which may be in the form of text, audio and video etc., and one file with having same name and same size or same folder with same size may occur on the storage device used as a cloud server. The present work is successfully removing the duplicate folders as well as duplicate files within a fraction of time. After removal of these the secure access is also established by the use of ElGamal and later on fuzzy concepts are introduced to reduce the response time for secure transmission of files from one end to another end. It is also observed that concept of fuzzy logic introduced in the ElGamal technique reduces the response time and increases the efficient transmission of unique files. The above data is considered only for four categories of files like .jpg, .pdf, .doc and .txt while it can be extended for other categories of files may also be considered for extension of present work. The average efficiency of data is also computed for these four files size over the ElGamal and fuzzy ElGamal algorithm. Since the text files take lesser size, therefore fuzzy ElGamal has 59.5% average efficient over the ElGamal algorithm used for secure transmission of data from one device to another device. For the file size related to .doc, .pdf and .jpg, the effective percentage of efficiency is 62.5%, 44.2%, 34.2% respectively. The above approach can be further extended by the use of Genetic algorithm as well DNA (De-oxyribo Nucleic Acid) security techniques.

References

[1] Seungmin Kang, Bharadwaj Veeravalli and Khin Mi Mi Aung, "A security-aware data placement mechanism for big data cloud storage systems", *IEEE 2nd international conference on big data security on cloud (BigDataSecurity), IEEE international conference on high performance and smart computing (HPSC) and IEEE international conference on intelligent data and security (IDS)*, IEEE, 2016. DOI: 10.1109/BigDataSecurity-HPSC-IDS.2016.

[2] ShankarNayak Bhukya, Suresh Pabboju and K Venkatesh Sharma, "Data Security in Cloud Computing Outsourced Database", *International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, pp.2458-2462.

[3] Daniya Rao, Deepak Painuli and Kamal Kant Verma, "Homomorphic Hybrid Encryption for Cloud Computing", International Journal of Advanced Research in Computer and Communication Engineering, Vol.5, No.5, pp.663-666, 2016.

[4] Arockia Panimalar.S, Subhashri.K, "Securing Outsourced Data on Cloud Using ElGamal Cryptosystem". International Research Journal of Engineering and Technology (IRJET), Vol. 4, pp.53-56, 2016.

[5] Keke Gai, Meikang Qiu and Hui Zhao, "Security-Aware Efficient Mass Distributed Storage Approach for Cloud Systems in Big Data", IEEE *2nd International Conference on Big Data Security on Cloud, IEEE International Conference on High Performance and Smart Computing, IEEE International Conference on Intelligent Data and Security*, pp.140-145, 2016. DOI: 10.1109/BigDataSecurity-HPSC-IDS.2016.68.

[6] Olugbenga W. Adejo, Isaiah Ewuzie, Abel Usoro, Thomas Connolly, "E-Learning to m-Learning: Framework for Data Protection and Security in Cloud Infrastructure", International Journal of Information Technology and Computer Science, Vol.10, No.4, pp.1-9, 2018.

[7] Preeti Soni, Rifaqat Ali, and Arup Kumar Pal, "A Two-factor based Remote User Authentication Scheme using ElGamal Cryptosystem", IoTSec'17 July 10-14, 2017, Chennai, India. pp.1-6, 2017. DOI: https://doi.org/10.1145/3084030.3084031

[8] Tao Xiang, Xiaogua Li, Fei Chen, Yuanyuan Yang and Shengyu Zhang, "Achieving verifiable, dynamic and efficient auditing for outsourced database in cloud", J. Parallel Distrib. Comput. Vol.112, No.1, pp.97–107, 2017. DOI: https://doi.org/10.1016/j.jpdc.2017.10.004.

[9] M. Thangavel, and P. Varalakshmi, "Enhanced DNA and ElGamal cryptosystem for secure data storage and retrieval in cloud". Cluster Computer, Vol. 21, pp.1411–1437, 2017. DOI: https://doi.org/10.1007/s10586-017-1368-4.

[10] D. Thiyagarajan, and G. R. Rao, "Ensuring Security for Data Storage in Cloud Computing using HECC- ElGamal Cryptosystem and GSO Optimization", International Journal of Intelligent Engineering and Systems, Vol.10, No.5, pp.115-124, 2017.

[11] P. Ravi Kumar, P. Herbert Rajb, and P. Jelciana, "Exploring Data Security Issues and Solutions in Cloud Computing". 6th International Conference on Smart Computing and Communications, ICSCC 2017, 7-8 December 2017, Kurukshetra, India. Procedia Computer Science (2018), Vol. 125, pp.691–697, 2017. DOI: https://doi.org/10.1016/j.procs.2017.12.089.

[12] Amjad Alsirhani, Peter Bodorik and Srinivas Sampalli. "Improving Database Security in Cloud Computing by Fragmentation of Data", *2017 International Conference on Computer and Applications (ICCA)*, IEEE, pp.43-49, 2017, DOI: 10.1109/COMAPP.2017.8079737.

[13] Navdeep Kaur, and Heena Wadhwa, "Security Enhancement in Cloud Storage using ARIA and Elgamal Algorithms", International Journal of Computer Applications, Vol. 171, No. 9, pp.19-23, 2017.

[14] Gagandeep Kaur, Mandeep Singh Devgan, "Data Deduplication Methods: A Review", International Journal of Information Technology and Computer Science, Vol.9, No.10, pp.29-36, 2017.

[15] Deepak Soni, Nishchol Mishra,"Multilevel Authentication based Data Security and Verification over Cloud Computing Environment", International Journal of Education and Management Engineering Vol.7, No.5, pp.56-68, 2017.

[16] M. Vedaraj, and Dr.M.Vigilson Prem, "A Hybrid Data Encryption Technique using two fish and Elgamal for cloud computing", International Journal of Management, Technology and Engineering, Vol. 8, pp.1473-1478, 2018.

[17] Yenumula Reddy. Big Data Security in Cloud Environment, *4th IEEE International Conference on Big Data Security on Cloud.* IEEE, pp.100-106, 2018.

[18] Alka Londhe, Vikrant Bhalerao, Suyog Ghodey, Sagar Kate, Niranjan Dandekar and Shubham Bhange, "Data Division and Replication Approach for Improving Security and Availability of Cloud Storage", *Fourth International Conference on Computing Communication Control and Automation*, pp.1-4, 2018, DOI: 10.1109/ICCUBEA.2018.8697355.

[19] Dhuratë Hyseni, Artan Luma, Besnik Selimi, and Betim Cico, "The Proposed Model to Increase Security of Sensitive Data in Cloud Computing", *International Journal of Advanced Computer Science and Applications*, Vol.9, No.2, pp.203-210, 2018.
[20] Youssef Gahi, Imane El Alaoui, "A Secure Multi-User Database-as-a-Service Approach for Cloud Computing Privacy", International Workshop on Emerging Networks and Communications (IWENC 2019) November 4-7, 2019, Coimbra, Portugal, Procedia Computer Science, Vol.160, pp.811–818, 2018. DOI: https://doi.org/10.1016/j.procs.2019.11.006.
[21] Fatma Mallouli, Aya Hellal, Nahla Sharief Saeed, Fatimah Abdulraheem Alzahrani, "A Survey on Cryptography: comparative study between RSA vs ECC Algorithms, and RSA vs El-Gamal Algorithms",*6th IEEE International Conference on Cyber Security and Cloud Computing (CS Cloud) 2019, 5th IEEE International Conference on Edge Computing and Scalable Cloud*, pp.173-176, 2019. DOI: 10.1109/CSCloud/EdgeCom.2019.00022.
[22] K. V. Pradeep, V. Vijayakumar, and V. Subramaniyaswamy, "An Efficient Framework for Sharing a File in a Secure Manner Using Asymmetric Key Distribution Management in Cloud Environment", Journal of Computer Networks and Communications, Vol. 2019, pp.1-8, 2019. DOI: https://doi.org/10.1155/2019/9852472.
[23] S.Ramamoorthy, B. Baranidharan, "CloudBC - A Secure Cloud Data access Management system", *3rd International Conference on and Communication Technologies*, pp. 217-220, 2019. DOI: 10.1109/ICCCT2.2019.8824828.
[24] Abdelrhman Sayed Awad, Adil Yousif and Gada Kadoda, "Enhanced Model for Cloud Data Security based on Searchable Encryption and Hybrid Fragmentation", *International Conference on Computer, Control, Electrical and Electronics Engineering*, pp.1-4, 2019. DOI: 10.1109/ICCCEEE46830.2019.9070918.
[25] Devi Thiyagarajan and Ganesan R. "Environmental Benefits of Enhanced Hecc - Elgamal Cryptosystem for Security in Cloud Data Storage Using Soft Computing Techniques". Foundation Environmental Protection and Research-FEPR. Ekoloji, Vol. 28, No.107, pp.665-677, 2019.
[26] Kaushik Sekaran, G.Raja Vikram, B.V. Chowdary, "Design of Effective Security Architecture for Mobile Cloud Computing to Prevent DDoS Attacks", International Journal of Wireless and Microwave Technologies, Vol.9, No.1, pp. 43-51, 2019.
[27] Priteshkumar Prajapati, Parth Shah, "A Review on Secure Data Deduplication: Cloud Storage Security Issue", Journal of King Saud University – Computer and Information Sciences, Vol.34, No.7, pp.1-12, 2020. DOI: https://doi.org/10.1016/j.jksuci.2020.10.021.
[28] Shailendra Kumar Tripathi, Bhupendra Gupta and K.K. Soundra Pandian, "An alternative practical public-key cryptosystems based on the Dependent RSA Discrete Logarithm Problems", Expert Systems with Applications, Vol.164, pp.1-13, 2020. DOI: https://doi.org/10.1016/j.eswa.2020.114047.
[29] P. Hima Bindu, T. Bhaskar Reddy, "An Exploration of Security Issues for Cloud Computing", Journal of Engineering Sciences, Vol. 11, No.1, pp.144-152. 2020.
[30] Kranthi Kumar Singamaneni and Sanyasi Naidu Pasala, "An improved dynamic polynomial integrity based QCP-ABE framework on Large cloud data security", International Journal of Knowledge-based and Intelligent Engineering Systems Vol. 24, No.2, pp. 145-156, 2020. DOI: 10.3233/KES-200037.
[31] Siddavaram Srivalli, P. Vijaya Raghavulu, "Encrypted Negative Password Using ELGAMAL", International Journal of Innovative Research in Technology, Vol.7, No.5, pp.117-123, 2020.
[32] Maya Mohan, M.K Kavithadevi and Jeevan Prakash V, "Improved ElGamal Cryptosystem for Secure Data Transfer in IoT Networks". *Proceedings of the Fourth International Conference on I-SMAC* (IoT in Social, Mobile, Analytics and Cloud). IEEE XplorePartNumber: CFP20OSV-ART pp.295-303, 2020. DOI: 10.1109/I-SMAC49090.2020.9243407.
[33] Vaishnavi Moorthy, Revathi Venkataraman, T. Rama Rao, "Security and privacy attacks during data communication in Software Defined Mobile Clouds", Computer Communications, Vol.153, pp.515–526. DOI: https://doi.org/10.1016/j.comcom.2020.02.030, 2020.
[34] Arfatul Mowla Shuvo, Md. Salauddin Amin and Promila Haque, "Storage Efficient Data Security Model for Distributed Cloud Storage", IEEE 8th R10 Humanitarian Technology Conference (R10-HT) pp.1-6, 2020. DOI:10.1109/R10-HTC49770.2020.9356962.
[35] Rida Qayyum, Hina Ejaz, "Data Security in Mobile Cloud Computing: A State of the Art Review", International Journal of Modern Education and Computer Science, Vol.12, No.2, pp. 30-35, 2020.
[36] Fursan Thabit, Sharaf Alhomdy, Sudhir Jagtap Dr, "Security analysis and performance evaluation of a new lightweight cryptographic algorithm for cloud computing", International Journal of Intelligent Networks Vol. 2, pp.18–33, 2021. DOI: https://doi.org/10.1016/j.gltp.2021.01.014.
[37] Sunitha Pachala, Ch. Rupa, and L. Sumalatha, "An improved security and privacy management system for data in multi-cloud environments using a hybrid approach", Evolutionary Intelligence, Vol.14, pp.1117-1133, 2021. DOI: https://doi.org/10.1007/s12065-020-00555-w.
[38] P. Zhang, H. Chi, J. Wang, and Y.F. Shang, "Data Security Protocol with Blind Factor in Cloud Environment", MDPI, Vol.12, No.9, pp.1-14, 2021. DOI: https://doi.org/10.3390/info12090340.
[39] Fursan Thabit, Sharaf Alhomdy and Sudhir Jagtap, "A new data security algorithm for the cloud computing based on genetics techniques and logical-mathematical functions", Global Transitions Proceedings, Vol.2, pp.18–33, 2021. DOI: https://doi.org/10.1016/j.ijin.2021.03.001.
[40] Amitesh Singh Rajput, Vishesh Kumar Tanwar and Balasubramanian Raman, "Z-Score-Based Secure Biomedical Model for Effective Skin Lesion Segmentation Over eHealth Cloud", ACM Trans. Multimedia Comput. Commun. Appl., Vol.17, No.2, pp.1-19, 2021. DOI: https://doi.org/10.1145/3430806.

Authors' Profiles

Deepika Gautam received Post Graduate Degree in Computer Applications (M.C.A.) from Dr. A.P.J. Abdul Kalam Technical University, Lucknow, Uttar Pradesh, India successfully completed various software projects and presently solving the problems related to security of cloud data as a research scholar in the Department of Computer Science from Babasaheb Bhimrao Ambedkar University, Lucknow, Uttar Pradesh, India under fellowship program of University Grant Commission(UGC), New Delhi. She has been interested in Cloud Computing, Cryptography and Security Protocols. Her research interest includes Cloud Data Security.

Prof. Suvendir Rimer received the B.Sc. degree in engineering (electrical) and the Higher Diploma degree in computer science from the University of the Witwatersrand, the master's degree in business administration from Bond University, and the master's degree in computer engineering and the Ph.D. degree in engineering from the University of Pretoria. Her academic experience includes stints as a Lecturer with the University of Pretoria and as a Senior Lecturer with the University of Johannesburg, South Africa. She has worked in industry as a Software Engineer and an Architect. She is currently an Associate Professor with the Department of Electrical and Electronic Engineering Science, University of Johannesburg. She is a Microsoft Certified Azure Solutions Architect Expert and an AWS Solutions Architect.

Prof. Vipin Saxena received his Ph.D. degree from Indian Institute of Technology, Roorkee, Uttarakhand, India. Presently, he is working as Professor in Department of Computer Science, Babasaheb Bhimrao Ambedkar University, Lucknow, India. He has published more than 210 research articles in the International and National Journals and Conferences, authored 05 books in the field of Computer Science and Scientific Computing, attended 59 International and National Conferences and received three National Awards for meritorious research work in the field of Computer Science. His research interests are Scientific Computing, Computer Networking and Software Engineering.

How to cite this paper: Deepika Gautam, Suvendir Rimer, Vipin Saxena, "Secure Access of Folders and Files after Removal of Duplicacy over the Cloud", International Journal of Computer Network and Information Security(IJCNIS), Vol.16, No.1, pp.48-60, 2024. DOI:10.5815/ijcnis.2024.01.04