Intelligent Automation & Soft Computing
DOI: 10.32604/iasc.2022.026049
Article

Data De-Duplication Process and Authentication Using ERCE with Poisson Filter in Cloud Data Storage

B. Venkatesan[1,*] and S. Chitra[2]

[1]Department of Information Technology, Paavai Engineering College, Namakkal, 637018, India
[2]Department of Computer Science and Engineering, Er. Perumal Manimekalai College of Engineering, Hosur, 635117, India
*Corresponding Author: B. Venkatesan. Email: bvenkatesh21@outlook.com
Received: 14 December 2021; Accepted: 09 February 2022

Abstract: The cloud storage is essential environment for users to access high confidential data. Every single data is most valued by users. If we count, day by day information as well as, memory storage are increasing gradually. Cost of memory will increase when data increases with demand for storage. At present data duplication or redundant data storing in the cloud became hard for storage provider. Also, it makes security issue if repeated data from various users stored in the server. It makes data duplication, which is very efficient for intruders. Also, when same data stored in cloud, there will be a waste of storage. Our research article is focused on security of original data by generating the key from owner and identifying the repeated data, while storing in cloud platform. This process is called as data de-duplication, which is also known as intelligent based computing. Storing the data in single instance is very challenging among cloud computing researchers. In this article we propose a content level de-duplication with re-encryption using enhanced Randomized convergent encryption (ERCE) based on Poisson filter (PF). First the data is encrypted and re-encrypted using the cipher methodology. Generated key only stored and processed by authenticated user. Owner of the data give permission to access key. Then the Poisson filter is used in de-duplication process. if key is authenticated, then the authenticated user can access data from cloud server. Data is stored only once and accessing key decides who can access the data. The result is evaluated with various existing algorithm. our proposed algorithm proves less time in downloading file and less computation cost when comparing with existing system.

Keywords: Cloud computing; data redundancy; de-duplication; enhanced randomized convergent encryption; Poisson filter

1 Introduction

In order to improve and utilize the cloud storage, data de-duplication is an important topic need to be considered to remove the redundant information in the system. The de-duplication is the process of which identify the multiple copy of the same data, remove the existing data/files from the storage and maintain the original copy stored. The data de-duplication needs are listed, (i) to reduce the cloud storage to

improve the efficiency of the cloud environment (ii) to provide accurate data retrieval and to reduce the communication cost. The de-duplication can be handled as block level and file level. Block level de-duplication is efficient compared to file level and it eliminates the repeated blocks by fixed size or variable size. File level de-duplication is simple as compared to block levels that eliminates the duplicate files by considering the minimum number of resources and reduce the amount of storage medium.

In the de-duplication process, the block of data are compared with the existing copy, if there is a match found, then the new block of data is replaced with the original data reference. So that the original copy of the data alone will be stored in the cloud, the repeated accessing of the same data need not be stored separately rather the original copy will be referred using the reference. The de-duplication process first proposed by [1], to decrease the redundant storage and bandwidth by removing the duplicate copies of that data and stored only one copy of that data. There are various de-duplication techniques are provided by the Cloud service providers such as Dropbox, Wuala, Mozy, and Google Drive [2].

The conventional encryption methods are used to encrypt and decrypt the data that are stored in the cloud using personnel key for secure data transfer. The randomized convergence encryption methods are developed for de-duplication with key management. This method not assures consistency and reliability. Secure de-duplication scheme is used to store the data and share the data in secure in cloud. In this method each owner of the data can generate a key and store it into the cloud for transfer. This will leads to high communication cost. To overcome these issues, an approach called enhanced secure content de-duplication identification and prevention (ESCDIP) [3] for de-duplication of file and content level. This method avoids the unauthorized access using secret key validation. Even though this technique provides secure data transfer, if there are more than one owner for same content means, then key for that content is confused. In paper [4] authors Proposed the content level de-duplication with Poisson filter that will allow random data in cloud transfer at minimum operation time.

1.1 Scope of the Work

- There are various algorithms are there in current scenario for de-duplication. But consistency with security and minimum communication time is still in research in cloud.
- The existing de-duplication algorithms are considering the dynamic changes of the owner using group key distribution. But data leakage is still an issue in the cloud to transfer the data to the user from the correct owner.

1.2 Objective of the Paper

- To overcome the issues like consistency, reliability, high communication cost and dynamic owner change, a novel enhanced encryption technique for de-duplication is need. With this in a mind, this work proposed an enhanced RCE with Poisson filter de-duplication re-encryption technique.
- This proposed methodology will optimize the memory utilization and remove the unwanted data storage in the cloud server that will leads to prevent the unauthorized access of the data in the cloud.
- The existing de-duplication algorithms are considering the dynamic changes of the owner using group key distribution. This proposed work focuses on the server side de-duplication, also it avoids the data leakage not only by users but also the previous owner of the same data.
- This proposed work enhances the data transfer between cloud and user with encryption based retrieval and also avoids the duplication using Poisson filter. Finally it manages the owner group using group key distribution.

Hence, this proposed work will be the solution to solve the consistency, reliability, communication cost, privacy and security in all in one method.

2 Literature Review

In paper [5] authors proposed the content—level de-duplication on streaming data using Poisson process filter technique at random level. This work used Poisson filter to choose the data that are already available or not by the classification approaches. These classification approaches is used to classify the streaming data into groups that perform the de-duplication on content level to store the cloud database. The semantic level of stream data with de-duplication is proposed in this paper to classify the type of data that are stored in the database.

In paper [6] authors proposed a server side de-duplication to encrypt the data in cloud. Randomized convergent encryption and group key distribution are the key approaches used to control the data access by cloud server. It obtains the data integrity and security. The comparative analysis of this work concludes that this work is better than other existing algorithms.

In paper [7] authors proposed the content level and file level de-duplication scheme with reliability on cloud. The user of the cloud have master key for encryption using the proposed technique that are stored in the cloud storage. This work identifies the duplicate data that are in the cloud. The experimented result of this work reduces the uploading and downloading time as 2.3 and 2.31 s compared to existing algorithms.

In paper [8] authors reviewed about the de-duplication mechanism with encryption on cloud storage system. To store the relevant data in the virtual environment the cloud storage space is compressed. In paper [9] authors proposed a de-duplication method called ZEUS to provide the privacy in cloud. This work minimizes the communication cost and it does not focus the similarity between the encrypted content in cloud. The authors in [10] proposed a predicate encryption de-duplication scheme in cloud. This will allow the file level de-duplication with the same user.

In paper [11] authors addresses the secure de-duplication data techniques by assume the data item based on the popularity of the data at different level. This technique used dynamic hash approach for de-duplication. In paper [12] authors proposed a novel de-duplication scheme based on biometrics in cloud. This approach created much bio-key generation based on fingerprint pattern extraction in order to improve the security. They experiment their work on AWS cloud services.

In paper [13] authors discussed about the study of de-duplication data techniques with performance metrics. The chunking techniques used in the de-duplication process. They compared different chunking algorithms. They summarized their work with the chunking algorithms advantages, disadvantages and research directions. In paper [14] authors provide a detailed review about the different techniques and methods for data de-duplication in cloud storage. They suggested that the data mining techniques can solve the de-duplication in cloud.

In paper [15] authors surveyed about the cloud computing threats, issues and solutions. They reviewed about the machine learning algorithms such as supervised, unsupervised and reinforcement learning to solve the security issues in the cloud. They compared and suggested their advantages, disadvantages with future directions on cloud storage research. In paper [16] authors proposed a security algorithm to improve the cloud system performance. Because of the data instability among the cloud user, there is leakage of data still exist. They used ANN for the de-duplication process.

In paper [17] authors reviewed about the issues and challenges in the cloud environment. In order to create the information flexibility and versatility, the Cloud computing is treated as an processing condition that can be registered and shared by the host over the cloud environment. They proposed a trust

based access control model to provide the security in distributed cloud environment. Based on the estimation of the trust, the client and cloud can access the data.

In paper [18] authors reviewed about the interruption recognition technique on mobile cloud that used computational perception. They depicted the cloud computing to mobile cloud computing diagram as a model. In paper [19] suggested the Machine learning based algorithms for data de-duplication and this is an under research to solve the best security de-duplication framework for cloud. Also the authors suggested to co-ordinate the heterogeneous LBS frameworks with cloud.

In paper [20] authors discussed that distributed denial of service can cause serious effect on the performance of the cloud. A DDoS assault is the major risk in security. It permits the assault as an interloper. There are various machine learning algorithms are used to identify the threats in the DDoS. The algorithm called C4.5 can be used for this based on decision tree can be used for order and grouping. While using c4.5, the decision rate can be achieved 98% than other algorithms. It provides the accurate result on both discrete and continuous data. The proposed a technique for de-duplication based on classification and they reported the advantages of the classification called IDPS over large data set. The disadvantages as memory consumption are stated. They used the K means & intrusion detection techniques for de-duplication with the objective to solve the challenges and successful security. These detection techniques obtain high data privacy, consistency and the information management is the disadvantage of these approaches.

3 Proposed ERCE-PF Data Deduplication

As large volume of data are stored in the cloud and shared between the users day by day cloud computing becomes a dominating research work to find the security in the cloud, and reducing the storage of the cloud. The storage management of the cloud is the major challenge since increasing the volume of data and also the same data can be stored in the cloud multiple times by the multiple users accessing the same data. To avoid the duplication entries of the same data in cloud will leads to provide the cloud service efficiently with less storage. Data de-duplication is the next research area in the cloud to solve the issues of repetition of the data over cloud. This will reduce the cloud server storage space. This proposed work called enhanced content de-duplication with randomized convergence encryption with Poisson filter is used to overcome the duplication issues in the cloud by preventing the unauthorized data access. The overview of the proposed work is stated in the following steps.

1. The owner of the data, store the file into the cloud once it is encrypted using ERCE. The existing de-duplication algorithms are considering the dynamic changes of the owner using group key distribution. This proposed work focuses on the server side de-duplication also to avoid the data leakage not only by users but also the previous owner of the same data.
2. The user of the cloud while trying to access the content, check for duplication using Poisson filter. If the content is already available then the owners are get notified.
3. The user request for the key to decrypt the data from the owner. Owner share the key to the user to get access the original content.
4. There is no duplicate copy of the content are stored. The original content itself accessed by the user with owner permission.
5. Hence the strong trust between the cloud user and cloud service provider retained.

The overall working principle of the proposed work is shown in Fig. 1. It consist of the data owner, cloud user, cloud service provider, and secure deduplication using ERCE-PF with dynamic owner change, cloud storage, authentication parts. The preprocessing of the implementation setup is described as follows.

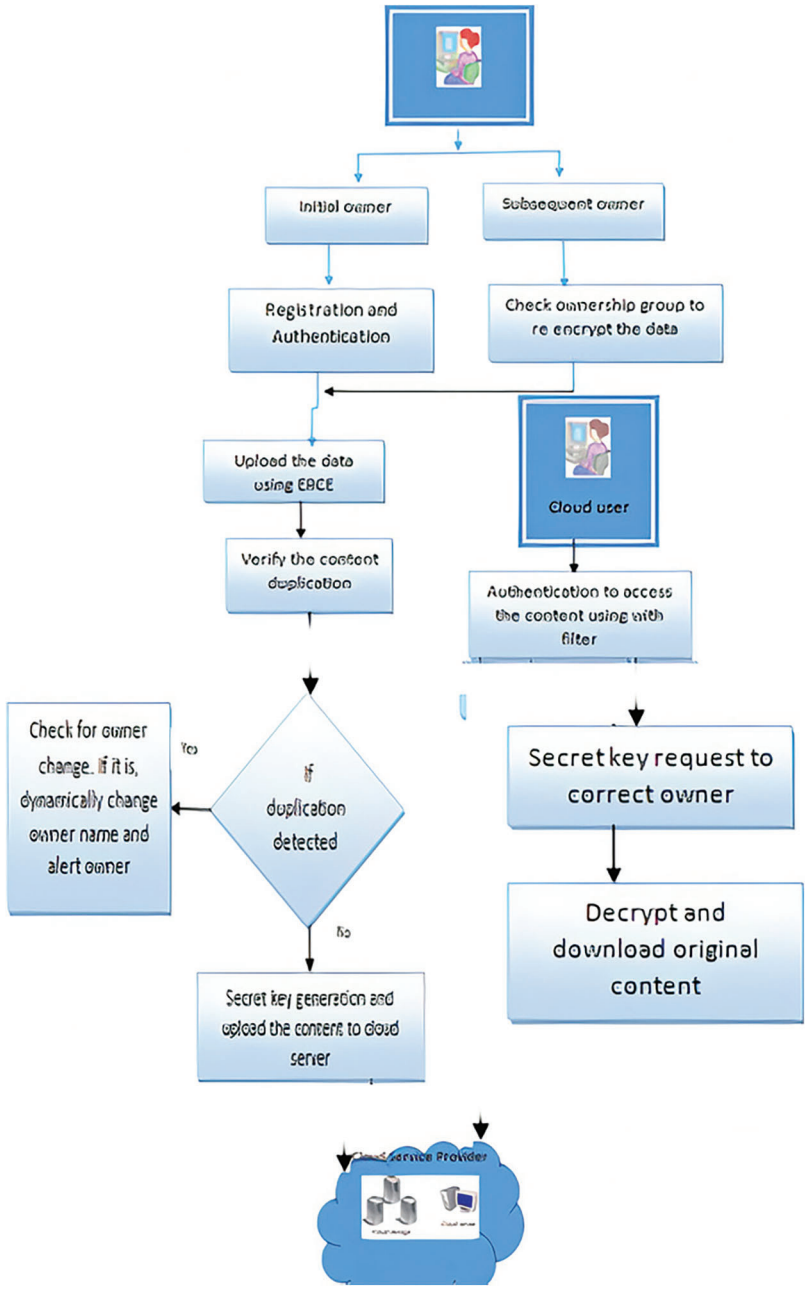

Figure 1: Working principle of proposed ERCE-PF architecture

3.1 Content Owner

The content owner is a client who is the owner of the data, who created the content and store it into the cloud to save cost with encryption. If the owner of the data who upload the content that is not already exists, then the owner is called as an initial owner, if the data already exists then the same content will be owned by existing owner, then the owner is called as a subsequent owner of that content. The same content shared by multiple owners is called as the ownership group. The current owner of the content will share the key to authenticated user for file sharing and he has the rights to delete the duplicated content to the user.

3.2 Cloud Service Provider

CSP provides the cloud storage and cloud service to the user Fig. 1. The cloud server deduplicates the data and then stores it into the cloud storage. It has the ownship records of the data with the tag, then manages the ownership lists using the group key to control the data access. When the user requests of the content, the server also checks for the deduplication task and then grant permission to access the content with the key distributed by the correct owner of the data.

3.3 Cloud User

The cloud user can access the requested content in the cloud by registering their general information and request for authentication. Validation checking is done by the cloud administrator to validate the user. Once validated, the key of the content will shared to download the requested content from the cloud. The available file accessed by the user after the owner grant permission with key.

3.4 Data Sharing and De-duplication

The requested data sharing will be done by dividing the requested data using proposed ERCE method that splits the data into equal fragments that are transformed into an understandable language. In this proposed work two kinds of encryption techniques are used i) data encryption with filter for dynamic owner change ii) data encryption with ERCE with PF for content de-duplication.

3.5 Proposed Algorithm: (Encryption using ERCE with Group Key Checking for Subsequent Owner and De Duplication with Poison Filter)

This proposed work focuses on content level de-duplication using encryption and filter techniques for single user and also manages the dynamic ownership if more than owner for same content. Initially the proposed work split the data into blocks and stores the data into the block. ERCE-PF performs the content level de-duplication using standard encryption technique to secure the content and find the duplicate copy of the data using poison filter. The data owner has the privileges to protect the data they uploaded using secret key, and he have the token for the content to identify the duplication. If the data owners are from the ownership group then the encrypted data will be again re encrypted using secret key and accessed from the cloud.

Algorithmic Process

Let U be the cloud users which contains $\{u_1, u_2 \ldots u_n\}$, M_i—data shared in the cloud, $G_i \in U$ is the ownership group. $L_i = \{T_i, G_i\}$—Ownership list for the data M_i in the cloud server. T_i—tag, KG_i—is the key for ownership key corresponds to each owner shared among G_i.

Step 1: Key Generation

KeyGen(U) = this take set of users U as input and generate key.

Step 2: Encryption

Encrypt(M,K) = encrypt the data with the key using enhanced randomized encryption algorithm which use hand based techniques and standard AES algorithm as $K_i = \mathrm{H}(M_i)$ and Tag $T_i = H(K_i)$ which generate the cipher text as C1 = $E_K(M_i)$ and C2 = K $\oplus$ K_i, $C_i = C1||C2$.

Step 3: Re encryption

Reencrypt(Cyphertext, G) = it takes the cypher text of the previous owner and check for the ownership group G. it re encrypt the cypher text so that the valid owner of the group will encrypt the data. Choose the random ownership key KG_i and re encrypt the data as $E_K\{KG_i(C_i)\}$ where K $\in$ $keygen(G_i)$.

(Continued)

Algorithmic: (continued).

Step 4: Deduplication using Poisson filter

Deduplicate(M) = Check the streamed data M that already exists or not using the tag generated for each owner of the data using Poisson filter based on classification algorithm such as SVM. It assigns the signature to the duplicated content for the identification. The duplicated data is represented as $DM = \{m_i, \{\sigma_i = \prod \sigma_{i,j}\}\}$ which is send to cloud server.

Step 4: Decryption

Decrypt(C,M,GK) = cloud user can decrypt the data M that encrypted as cypertext C data with the key from the correct owner in GK. The user u decrypt the data as $C_i = D(KG_i)$, L = $C_i \oplus K_i$, $M_i = D_L(C_i)$ and Tag $T_i = H(M_i)$

Pseudo code

Input: Data content M (txt,pdf and doc), User U, ownership group G,

Output: avoid the data duplication and allow privileged access of the content

Start

1. Process the data owner authentication and generate key for the data M.

2. If(owner = initial owner) then

3. Registration and generate key for the data M using the GenKey method and encrypt

4. the data using ERCE method. Uploaded the encrypted data with the proposed ERCE to the cloud server as upload($T_i||C_i$). Now the cloud server create ID for the owner in the group as IDG_t and insert the record.

5. Else (subsequent owner)

6. Check the Group key function and re encrypt the data using re encrypt method

7. End if

8. If duplication detected using Poisson filter

$$P[M(t) = m) = \frac{e^{-\lambda t}(\lambda t)^m}{m!}$$

where, t ≥ 0 and m = 0,1,2,3…n, e = constant approximately 2.71 and λ = parameter that find the record with good time interval

Else

Upload the data to cloud with ERCE method

9. Cloud user authentication process

10. Request for the content with user registration

11. Owner of the data receive request and validated

12. Valid user can get the token from the server

13. If token = valid

Download the original content with the shared key

Else

Block the user

Failed to download

14. End if

15 End

Hence the proposed algorithm with enhanced randomize convergence encryption based on hash based technique with Poisson filter is efficient method for deduplication of the content that are shared in the cloud. With the innovation of checking the owner as an initial owner or subsequent owner with the group list will enhance the work more secure and utilize the cloud server storage in an efficient way. This will improve the data uploading and downloading time as fast as the existing algorithms and make the cloud environment as a secure one.

4 Experimental Results and Discussions

The cloud setup with the proposed work is implemented in Java programming language by using the packages called java.security and javax.crypto. The proposed work is evaluated using the data files of different sizes which is represented in Tab. 1.

Table 1: Evaluation data settings

S.No	File Type	Size (MB)
1	Text	1
2	Document	2
3	PDF	4
4	Image	5
5	Music	10

4.1 Evaluation Criteria

This proposed algorithm is evaluated under the simulation setup based on the execution time, file uploading time and file downloading time. The efficiency of the proposed work is analyzed by compared with the existing algorithms on encrypted data in terms of theoretical and practical aspects of the communication cost.

i) Encryption and Decryption Time: To evaluate the proposed ERCE scheme, the encryption and decryption time are calculated and shown in Tab. 2 which is depicted in Fig. 2.

Table 2: Encryption and decryption time of proposed ERCE

Data size	Time (ms)	
	Encryption time	Decryption time
1	4	4
2	6	6
4	12	11
5	20	18
10	35	33

Fig. 2 shows the time taken for encryption and decryption process of our proposed work. As the data size increases the time proportionally increases. Hence the encryption and decryption process time is directly proportional to the size of the data.

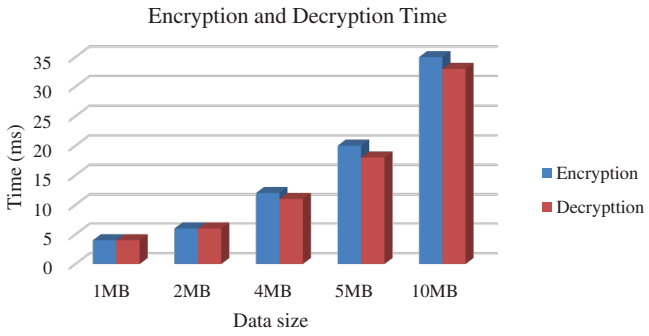

Figure 2: Encryption and decryption time of ERCE

ii) File Uploading Time: it is the total time of encryption process and data owner's contribution time in the cloud.

$$FUT = encryption\ time + \{Endtime - StartTime\}$$

FUT-File upload time, Endtime is the file uploading completion time and start time is the initial start time to upload the file.

iii) File Downloading Time: it is the total time of data decryption process and data downloading time form the server.

$$FDT = \frac{(T_{end} - T_{Process})}{ONCSP_{bandw}} + T_{decrypt}$$

FDT-File download time, T_{end} is the time taken to download the content, $T_{Process}$ is the processing time of the view of the content, $T_{decrypt}$ is the time taken to decrypt and view the original file and $ONCSP_{bandw}$ is the cloud service provider bandwidth selected by the owner.

iv) Total Communication Cost: it is the estimation of the total transmission time with respect to the data size.

$$TCC = \frac{Data\ transfer\ rate}{data\ size} \times 100$$

Tab. 3 shows the file upload time in milliseconds, file download time in milliseconds and total communication cost in percentage.

Tab. 3 represented the file upload time, download time and communication cost for the proposed enhanced RCE using Poisson filter de-duplication with efficient re-encryption. Which is depicted in Figs. 3–5.

Figs. 3–5 shows the file uploading, file downloading time and communication cost with respect to the different data sizes 1 MB, 2 MB, 4 MB, 5 MB and 10 MB. Hence the integrated proposed method with ERCE-PF with efficient re-encryption shows better result in terms of minimum uploading, downloading time and reducing communication cost. For evaluating the performance of the proposed algorithm, it is compare with the existing algorithms such as convergent encryption (CE), leakage resilient (LR), randomized convergent encryption (RCE), secure de-duplication scheme (SDS) and ESCDIP.

Tab. 3 shows the performance evaluation of proposed scheme with existing algorithms in terms of FUT, FDT and TCC. Fig. 6 shows the performance evaluation of file uploading time of various algorithms.

Table 3: Evaluation parameters of proposed algorithm

Evaluation Parameters	Size of the Data				
	1 MB	2 MB	4 MB	5 MB	10 MB
FUT (ms)	1.34	2.67	3.89	8.675	11.532
FDT (ms)	1.32	2.54	3.45	8.433	11.211
TCC (%)	110	100	95	92	85

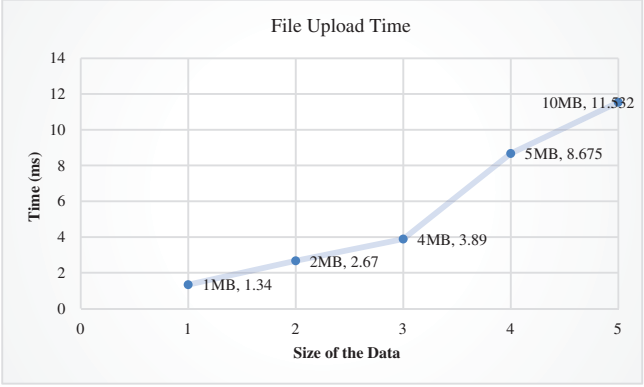

Figure 3: File upload time of proposed scheme

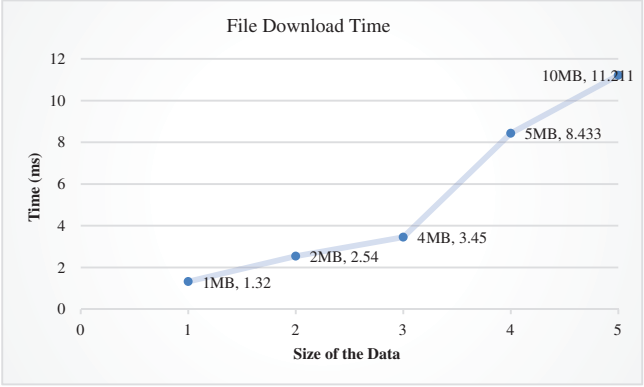

Figure 4: File download time of proposed scheme

From all the evaluation experiments, our proposed work shows much better result in terms of encryption time, decryption time, File uploading time, File downloading time and Total communication cost compared to other existing algorithms. The next to our proposed method is ESCDIP. The RCE are good de-duplication algorithm but lack in consistency and reliability. Our proposed work is efficient in terms of security and also it save the cloud storage efficiently compare to other existing algorithms in Figs. 7 and 8.

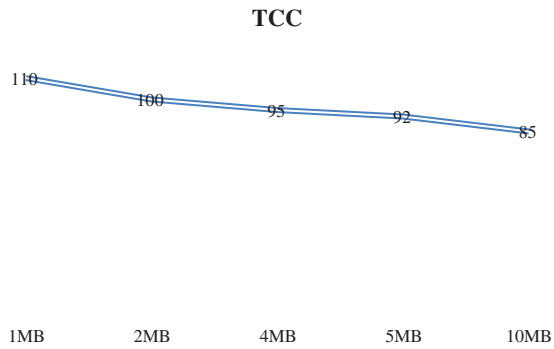

Figure 5: Total communication cost of the proposed scheme

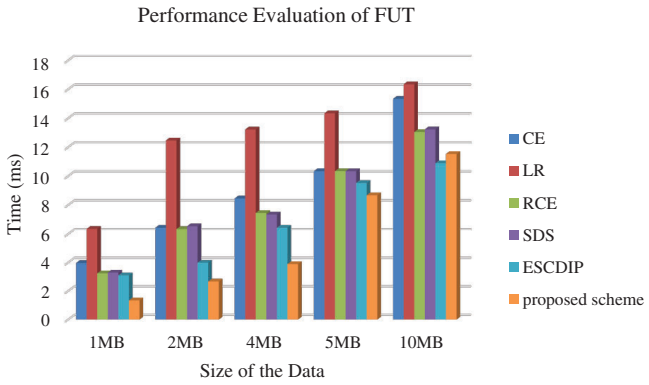

Figure 6: FUT comparison of various de-duplication algorithms

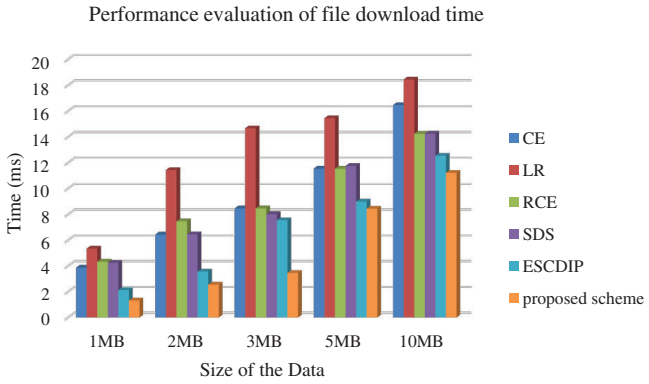

Figure 7: FDT comparison of various de-duplication algorithms

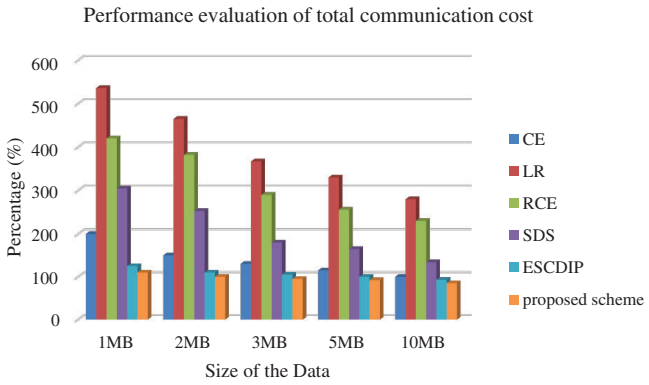

Figure 8: TCC comparison of various de-duplication algorithms

5 Conclusion

Data redundancy and storing redundant data in cloud is very serious problem. Our article tackles this problem by proposing ERCE-PF in saving the instant data in database without duplication. The user authentication process is assured by generating key. Owner of the data decides whom to access the data. The encryption process is done twice. This process checks the originality of the authenticated user and processed using de-duplication algorithm. Further Poisson filter helps for better elimination of the redundancy in the experiment. Main advantage of this experiment is, data stored only once. User who tries to access data needs key and access permission from owner. So we store data only in single location in server and user authentication is ensured by algorithm. The result is evaluated using various existing de-duplication algorithm. Our proposed result takes less time for execution. Previously RCE is considered as lack of reliability and execution performance. Enhanced RCE is overcomes the drawbacks and improves the performance of the system. In future, the hybrid filter techniques with de-duplication help to improve the performance by eliminating repeated data.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] W. J. Bolosky, S. Corbin, D. Goebel and J. R. Douceur, "Single instance storage in Windows 2000," in *Proc. Conf. on Usenix Windows Systems Symp.*, Berkeley, CA, 2000.

[2] J. Hur, D. Koo, Y. Shin and K. Kang, "Secure data deduplication with dynamic ownership management in cloud storage," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 11, pp. 3113–3125, 2016.

[3] J. K. Periasamy and B. Latha, "An enhanced secure content de-duplication identification and prevention (ESCDIP) algorithm in cloud environment," *Neural Computing & Applications*, vol. 32, no. 2, pp. 485–494, 2020.

[4] G. U. Devi and G. Supriya, "Encryption of big data in cloud using de-duplication technique," *Research Journal of Pharmaceutical, Biological and Chemical Sciences*, vol. 8, no. 3, pp. 1103–1108, 2017.

[5] C. M. Yu, S. P. Gochhayat and M. Conti, "Privacy aware data deduplication for side channel in cloud storage," *IEEE Transactions on Cloud Computing*, vol. 17, no. 1, pp. 597–609, 2018.

[6] J. Li, X. Chen, X. Huang, S. Tang, Y. Xiang *et al.,* "Secure distributed deduplication systems with improved reliability," *IEEE Transactions on Computers*, vol. 64, no. 2, pp. 3569–3579, 2015.

[7] Y. Shin and K. Kim, "Equality predicate encryption for secure data deduplication," in *Proc. Conf. on Information Security and Cryptology (CISC-W)*, Beijing, China, pp. 64–70, 2012.

[8] R. Tirapathi, M. V. P. Burramukku and C. Sekhara Rao, "Data deduplication in cloud storage using dynamic perfect hash functions," *Journal of Advanced Research in Dynamical and Control Systems*, vol. 9, no. 12, pp. 2121–2132, 2017.

[9] R. Ramya Kalangi and R. Chandra Sekhara, "A novel cloud user to service authentication framework using biometrics," *International Journal of Advanced Science and Technology*, vol. 134, pp. 19–32, 2020.

[10] A. Bhalerao and A. Pawar, "A survey: On data deduplication for efficiently utilizing cloud storage for big data backups," in *Proc. Int. Conf. on Trends in Electronics and Informatics (ICEI)*, Tirunelveli, pp. 933–938, 2017.

[11] S. Lalitha and N. Kamal Raj, "A survey on data de-duplication methods in cloud storage system," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 5, no. 7, pp. 1–12, 2017.

[12] U. A. Butt, M. Mehmood, S. B. Shah, R. Amin, M. W. Shaukat *et al.,* "Review of machine learning algorithms for cloud computing security," *Electronics*, vol. 9, pp. 1379–1392, 2020.

[13] A. N. Khan, M. Y. Fan, A. Malik and R. A. Memon, "Learning from privacy preserved encrypted data on cloud through supervised and unsupervised machine learning," in *Proc. Int. Conf. on Computing, Mathematics and Engineering Technologies*, Sindh, Pakistan, pp. 1–5, 2019.

[14] S. Shamshirband, M. Fathi, A. T. Chronopoulos, A. Montieri and F. Palumbo, "Computational intelligence intrusion detection techniques in mobile cloud computing environments: Review, taxonomy, and open research issues," *Journal of Information Security and Applications*, vol. 55, pp. 1–52, 2019.

[15] Z. Li, K. Xu, H. Wang, Y. Z. Wang and X. Shen, "Machine learning based positioning: A survey and future directions," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, pp. 96–101, 2019.

[16] A. Meryem, D. Samira and E. O. Bouabid, "Enhancing cloud security using advanced map reduce k-means on log files," in *Proc. Int. Conf. on Software Engineering and Information Management*, New York, NY, USA, pp. 63–67, 2018.

[17] J. Chen, L. Liu, R. Chen and W. Peng, "Secure outsourcing of high-order singular value decomposition," in *Proc. Australasian Conf. on Information Security and Privacy*, Wollongong, Australia, pp. 309–329, 2020.

[18] V. Sheng and J. Zhang, "Machine learning with crowdsourcing: A brief summary of the past research and future directions," in *Proc. AAAI Conf. on Artificial Intelligence*, Honolulu, HI, USA, 2019.

[19] Y. Zhao, J. Chen, D. Wu, J. Teng and S. Yu, "Multi-task network anomaly detection using federated learning," in *Proc. Tenth Int. Symp. on Information and Communication Technology*, Island, Korea, pp. 273–279, 2019.

[20] U. A. Butt, M. Mehmood, S. B. Shah, R. Amin, M. W. Shaukat *et al.,* "A review of machine learning algorithms for cloud computing security," *Electronics*, vol. 9, no. 9, pp. 1379, 2020.