ISSN: 2395-7639

International Journal of Multidisciplinary Research in Science, Engineering, Technology & Management (IJMRSETM)

(A Monthly, Peer Reviewed Online Journal)

Visit: www.ijmrsetm.com

Volume 7, Issue 5, May 2020

Cloud Storage for Secure Data Uploading With Chunk-Based Deduplication

Mrs.S.Ruba, Dr.A.M.Kalpana

Assistant Professor, Department Computer Science and Engineering, Government College of Engineering, Salem(D.t), India

Professor/Department of Computer Science and Engineering,Government College of Engineering, Salem(D.t), India

ABSTRACT: Cloud computing was based on the infrastructural and conceptual foundations of secure computing. Cloud-based services and service providers are being developed, which has led to a cloud-based technology trend. In the cloud environment, the privacy protection mechanism is based on the cutting process and deduplication analysis with privacy protection in data storage. If data security is not strong and consistent, the flexibility and benefits of deep learning must ensure reliability. In the proposed system, the user uploads new data to cloud computing. Two processes are applied in security analysis: duplication and fragmentation. The keyword of the user's question, which is compared with the reserve index of the hash term, and this term is saved in the cutting process. The proposed Reliability Based Neural Cooperative Filter (RNCF) algorithm is used to block cloud data content in three ways such as accuracy, application point and application time. RCNF algorithm. The reserve index term was compared with deduplication analysis. If partial information does not match the user's query, the query is answered for the user. It is assumed that if the partial information matches the user's keyword, the proposed deep learning (DL) algorithm is used for security reasons. The security process uses the Distributed Secure Cloud Data Chunk (DSCDC) algorithm for user authorization. If the thread is parsed, it will block the user. Deduplication is analyzed and the security level is verified using the proposed DL algorithm. Experimental results on a real dataset of cloud data prove its effectiveness, security and efficiency.

KEYWORDS: Deep Learning, Cloud Computing, Deduplication, Chunking, Security, Filtering algorithm and DSCDC.

I. INTRODUCTION

Infrastructure management is one of the most important challenges for companies/organizations that need to deploy large data sets and unified services using multiple local resources. These developments themselves are not affordable and very expensive for most organizations. On the other hand, it is essential in terms of usability security and data reliability, which are the main pre-given needs for the implementation of the company's strategy. All these features accelerate the use of cloud services for business data processing.

The main contributions of the proposed scheme are fourfold:

- A new Distributed Secure Cloud Data Chunk (DSCDC) algorithm with authorized duplication in cloud services was proposed. The algorithm is based on protecting data privacy and achieving secure data dissemination, uses role re-encryption for authorization.
- Secure data replication uses a deep learning algorithm that verifies the ownership of the authorized user. RNCF is implemented to make the data slicing process more efficient. Return of ownership is ensured by the proposed method.
- Data upload and data security analysis are essential for the proposed system. The algorithm is protected according to the proposed security model, and the performance evaluation determines the effectiveness and efficiency of the proposed system.
- With the rapid growth of cloud computing efficiency in computing, the cloud server provides efficient storage space and scalable computing to users anywhere and anytime.

ISSN: 2395-7639

International Journal of Multidisciplinary Research in Science, Engineering, Technology & Management (IJMRSETM)

(A Monthly, Peer Reviewed Online Journal)

Visit: www.ijmrsetm.com

Volume 7, Issue 5, May 2020

Along with the increasing growth of cloud data, there is massive duplicate data was occupied in the cloud storage spaces and bring severe challenges to the limited cloud storage spaces. To tackle these issues this research proposed a Deep Learning (DL) algorithm for authorized user analysis through the deduplication process. Distributed Secure Cloud Data Chunk (DSCDC) algorithm is a role for novel secure authorized users, is analyzed in this approach. The deep learning method is used for security data analysis for processing storage, which was the main issue from the cloud data replication from multiple users. The research analyzes the duplicated data while uploading data from users through the hash index term value by the chunking method.

II. RELATED WORK

EmnaBaccour et al [1] Collaborating Mobile Edge Computing (MEC) servers are an emerging paradigm using video caching where cloud services are extended to edge networks to allocate multimedia content resources to end users. Through Content Delivery Networks (CDN), despite the fact that traffic is minimized during peak hours due to high demand. Exploiting cellular frequencies, device-to-device (D2D) communication has proven to be very efficient and burdensome for user devices. The research proposes a Collaborative Edge (CE) network with the CE-D2D framework to solve video cache maximization with efficient cellular network and bandwidth. In the proposed system, instead of one edge node downloading and caching videos, only partial video viewing is cached. Linear CE-D2D framework program according to schedule and cooperation between them was smooth with available resources.

Aobing Sun et al [2]Cloud computing, which used data or applications and stored data through the Internet, can be done in three ways: public cloud, private cloud or hybrid cloud. However, cloud users lack effective information security that has measurably advanced in understanding the security situation of their information infrastructure. In this literature, security evaluation consisting of an API from various cloud evaluations including video search, security recovery, scan engine, quantifiable evaluation, etc. With the help of computers, the security model consisted of a set of assessment elements consistent with their various. in areas such as storage, computing, application security, maintenance, networking, etc. Each item has three times reserved for vulnerabilities, points and repair method. Implement measurable evaluation based on the G-Cloud platform for different cloud users. It showed the dynamic security inspection score of one or more cloud users with visual charts and guided users to change the structure, improve the procedure and fix vulnerabilities to improve the security of cloud resources.

Wenhai Sun et al [3] a backup storage company investigated a secure track-based replication system in research. The author proposed to generate a random key based on these internal backup services. In the current committed work, one client finally exposed a client storage scheme that provided intuitive functionality to achieve multi-client protection with minimal performance loss. In addition to each backup, there was a rate-limiting policy that slowed down the brute force attack method. The proposed method in the malicious model, which is provably secure, also adapts the system design to take into account practical deduplication needs to achieve comparable plaintext deduplication efficiency.
Xian Weiquan et al [4] Cloud computing and the security of public cloud data was the biggest challenge of the research work. Therefore, this research model responds to the public cloud thread analysis and presents information security, information management and data center software / hardware and controls the intensity of migration. The balance between the risk factor and business life is a difficult aspect of this study. Summary of this cloud computing information security modeling risk prevention paper. Manage the global core defense of information security using various strategies using security access control.

III. PROPOSED METHODOLOGY

A) CLOUD COMPUTING

Big Sensing Data arises both in industrial applications and in scientific research, where data is produced at high volume and speed. Cloud computing offers a promising platform for processing and storing large measurement data because it provides a flexible stack of massive computing, storage and software services in a scalable way. Detection of big data in the cloud has adopted some techniques such as data compression. However, due to the large volume and speed of big data detection, the traditional data compression did not have enough scalability and efficient process.

ISSN: 2395-7639

International Journal of Multidisciplinary Research in Science, Engineering, Technology & Management (IJMRSETM)

(A Monthly, Peer Reviewed Online Journal)

Visit: www.ijmrsetm.com

Volume 7, Issue 5, May 2020

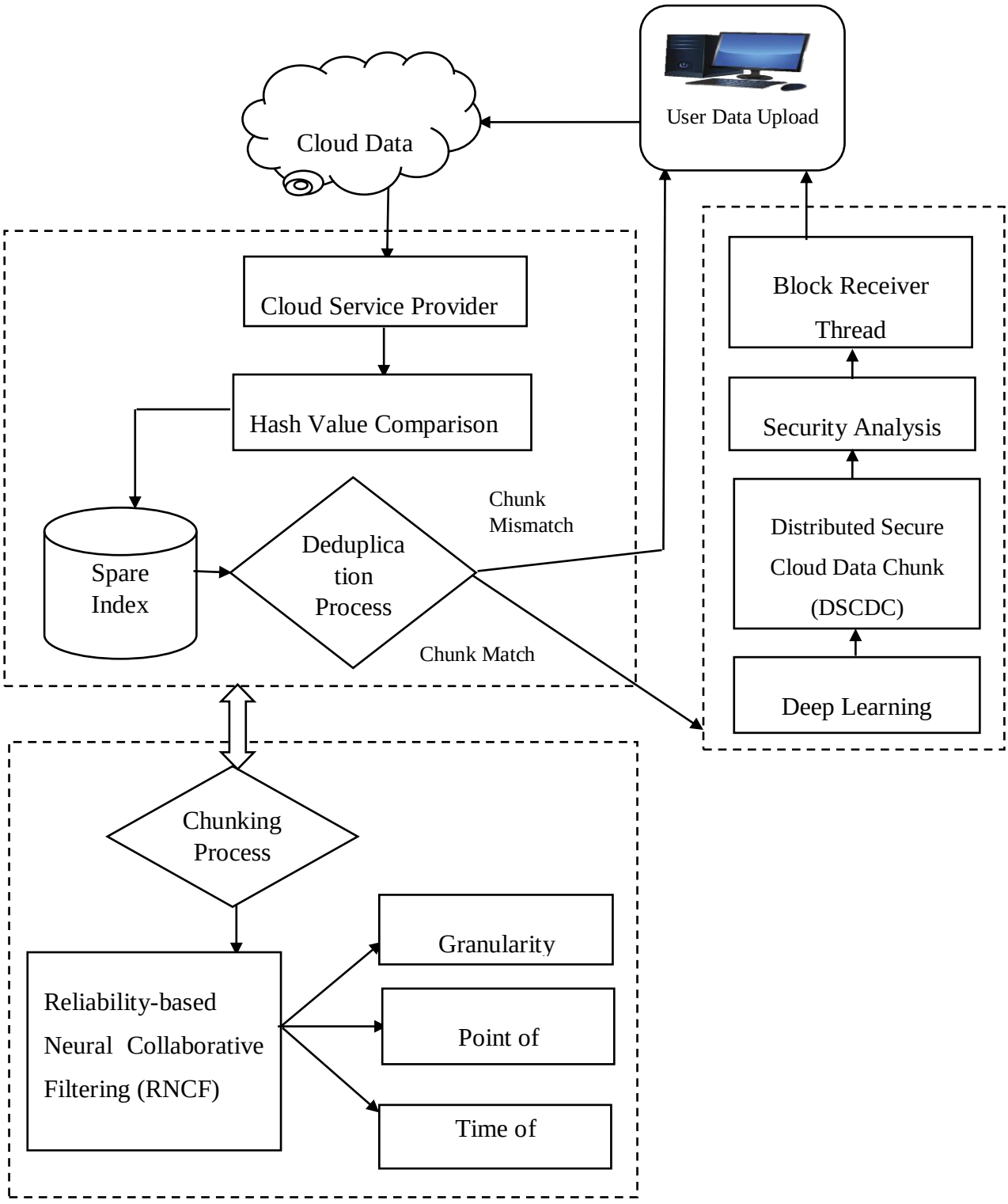


Figure 1: Overall Data Flow Process

ISSN: 2395-7639

International Journal of Multidisciplinary Research in Science, Engineering, Technology & Management (IJMRSETM)

(A Monthly, Peer Reviewed Online Journal)

Visit: www.ijmrsetm.com

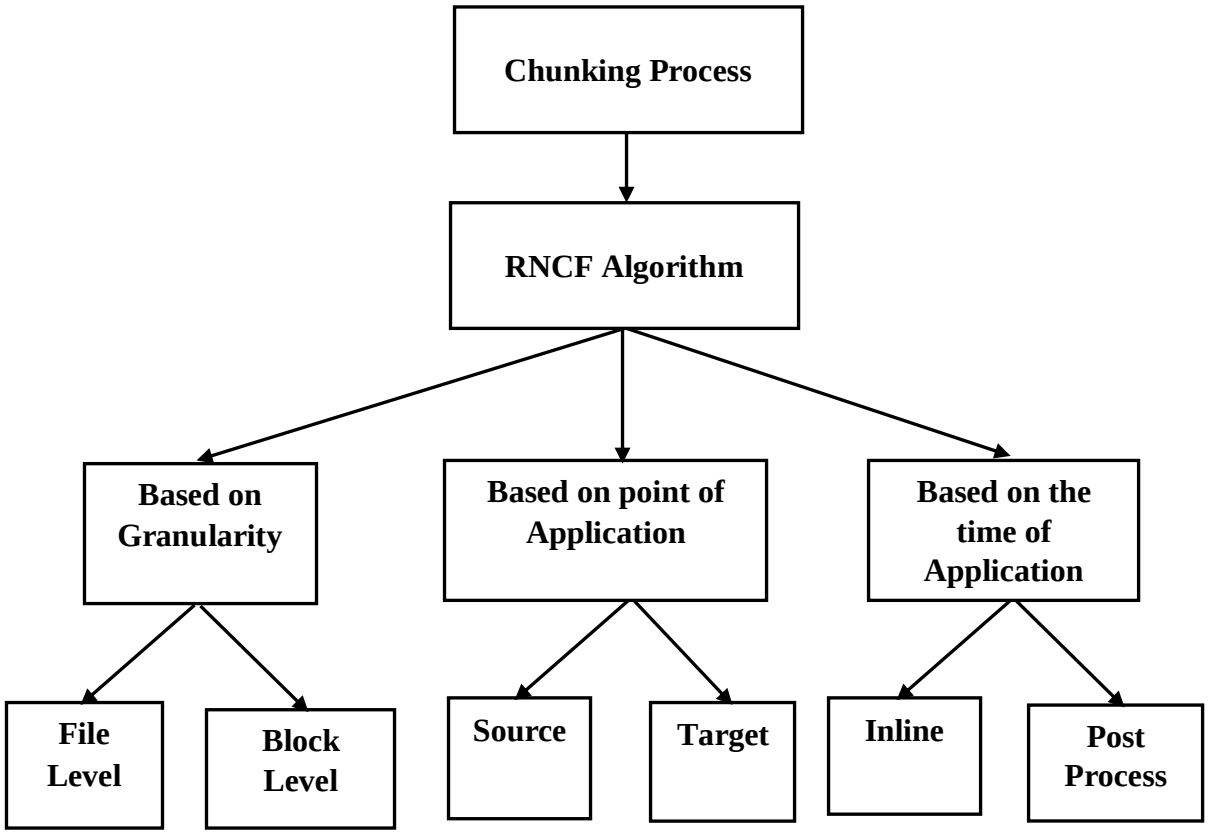
Volume 7, Issue 5, May 2020

Based ona certain data compression method, the author proposed scalable data by measuring the similarity of the sub-method data part. Instead of data compression techniques based on data units, chunk techniques are used in the proposed system. A deep learning (DL) algorithm is used to increase the scalability of cloud data. It has been experimentally shown that large measurement data using the deduplication similarity index improves efficient scalable compression techniques with little loss of data accuracy.
If the user uploads data to the cloud platform, the cloud service provider must enable the upload and use of the data. For the data security process, the proposed system uses data multiplication techniques to avoid replication. Double checking of data was analyzed by DSCDC algorithm and chunking was done by RNCF algorithm to filter by distraction of the cloud data to facilitate the analysis.

B) Reliability-based Neural Collaborative Filtering (RNCF)
The proposed architecture (RNCF) focuses on deeplearning to discover complex relationships to predict classification values and reliability. Extensive deep learning feature filtering technology is applied to the cloud platform to improve overall results. The cloud data was filtered by a partitioning method using RNCF techniques to partition the data as shown in the equation below.

- y = (y11,...,ydn) where ybj represents the vacant capacity of storage j and ‘b’denotes to the user,
- C = c1,...,cn ( ) where ‘cj’states cost of transmitting chunks to the storage j,
- z = (z11,...,zdn) where zbj denotes the ‘j’ storage of the b user.

System Characterization to describe the problem, it is assumed that the system consists of different users. Each user has different warehouses in different regions, and the total number of warehouses is indicated by n. It was previously mentioned that the transfer object is divided into small parts and these parts must be stored in the cloud storage. Clouds are described by the following vectors:



Figures 2: Cloud Chunk Data Processing

ISSN: 2395-7639

International Journal of Multidisciplinary Research in Science, Engineering, Technology & Management (IJMRSETM)

(A Monthly, Peer Reviewed Online Journal)

Visit: www.ijmrsetm.com

Volume 7, Issue 5, May 2020

Assume that the source chunks are stored in distributed cloud storages.The chunk transmission from the source location to the target storage location is characterized by a binary matrix 'X', where the elements are xij = 0,1 { } and

$$\sum_{i=}^{n} x_{ij} = 1, i = 1,...., m \quad (1)$$

Thus, the management cost is minimized and the quality of efficient storage issues are improved in a cloud service provider. In other words, if the chunk is transferred to the storage, the chunk marks this as a transferred one.

$$\sum_{i=}^{n} x_{ij} a_i \le y_{bj}, k \le m, b = 1,..., d, \; j = 1,...., n \quad (2)$$

$$\sum_{i=1}^{m} \sum_{i=}^{n} x_{ij} a_i \le \sum_{j=1}^{n} y_{bj}, b = 1,\ldots, d \quad (3)$$

Equation (3) describes that the size of the chunks which is uploaded to given storage cannot exceed the storage size. At the same time, condition represents that the overall size of the chunks in which user data uploaded to the cloud storage that must be less than the overall cloud storage size.

Chunking and Deduplication Process	Description
File Level Deduplication	File found unique then it is stored in the index table, but if the file is not unique then a specific file is stored.
Block Level Deduplication	Files are generally broken down named as a chunk and checkingfor redundancy this approach as per the name sub-files are operated.
Source-Based Deduplication	To the cloud non-duplicated data is backed up it is helpful for optimization and better utilization of the resources. It helps to increase backup by new users.
Target Based Deduplication	When the user upload/request data was not presented in the location, then this approach removes redundancy which means the deduplication process occurred and then data has been stored.

ISSN: 2395-7639

International Journal of Multidisciplinary Research in Science, Engineering, Technology & Management (IJMRSETM)

(A Monthly, Peer Reviewed Online Journal)

Visit: www.ijmrsetm.com

Volume 7, Issue 5, May 2020

The inline deduplication	This approach doesn't have any knowledge about files. It checks the raw block of the incoming file. This force extended of deduplication process is less and fixed length of the block only applied for deduplication.
Post Process Deduplication	After data stored in the device, the deduplication was checked. It can be applied both whole-file and sub-file. So the file data checksum can easily be compared with the previous entire backed-up file to eliminate deduplication easily.

Table 1: Chunking and Deduplication of Cloud Data

C) SECURE DATA DEDUPLICATION

Secure data deduplication techniques are used to analyze and remove redundant data, and it has been widely used in cloud storage to reduce storage space and download high bandwidth. When data is loaded, a hash value is generated and compared with the hash value of the existing stored backup index, if no duplicate value is found, the user has uploaded the file to the server of the cloud service provider. Assume that if a duplicate file is found, the process continues with deep learning methods for security analysis. The proposed Distributed Secure Cloud Data Chunk (DSCDC) algorithm is used for security analysis and user blocking.

The main goal of the research there is

- No redundancy of data
- Minimum storage space
- Efficient security process.

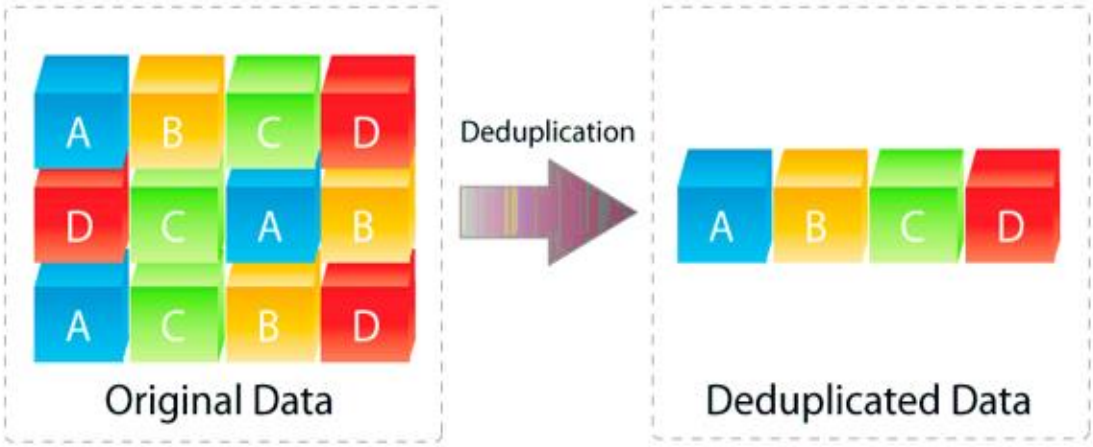


Figure 3: Analyzing and removing Deduplication

However, there are large data file was uploaded by the user's then it difficult to access the file for redundancy process. So, the data compression techniques were identified by Intra, and interfile duplicates checked. In this proposed research, the chunking process is applied to minimize the file into block-level and separated by granularity, point of application, and time of application. The higher deduplication ratio dr= original size of dataset stored, or space reduction percentages sr=1-1/dr.

In this research, a secure chunk deduplication system is applied to the storage device so that users can safely download files. If the hash value given to the user was compared with an existing hash value, it is matched with an existing stored

ISSN: 2395-7639

International Journal of Multidisciplinary Research in Science, Engineering, Technology & Management (IJMRSETM)

(A Monthly, Peer Reviewed Online Journal)

Visit: www.ijmrsetm.com

Volume 7, Issue 5, May 2020

file, then the distributed secure cloud data piece (DSCDC) deep learning algorithm authorizes the user and denies users when the thread is analyzed.

IV. RESULT AND DISCUSSION

Performance Analysis

In the proposed system, deduplication approach uses chunking by various level and the efficiency, storage utilization, throughput, bandwidth utilization, cost and deduplication ratio was compared and analyzed in the below table.

Deduplication Approach	Storage Utilization	Bandwidth Utilization	Efficiency	Throughput	Cost	Deduplication Ratio
File Level	Medium	Low	Low	High	Low	Low
Block Level	High	Medium	High	Low	Medium	High
Source Level	Medium	Low	Medium	Low	Low	Medium
Target Level	High	Low	Medium	Medium	High	Medium
Inline Process	Low	Low	Medium	Low	Medium	Low
Post Process	Low	High	High	Medium	High	High

Table 2: Comparison of various Data Deduplication Approaches

The storage process was differentiated with the various process, storage without deduplication, storage with deduplication and security level of deduplication were analyzed in the below comparison chart.

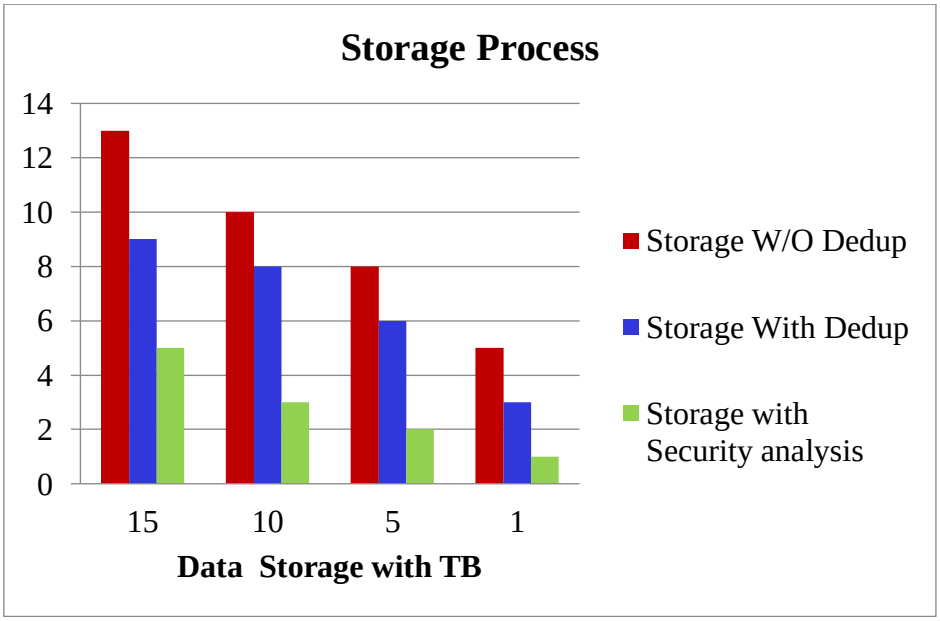


Figure 4: Cloud Data Storage comparison

After applying the security process, the duplication data was reduced and the figure demonstrated the security applied data which was taken a minimum amount of storage space in the proposed system

ISSN: 2395-7639

International Journal of Multidisciplinary Research in Science, Engineering, Technology & Management (IJMRSETM)

(A Monthly, Peer Reviewed Online Journal)

Visit: www.ijmrsetm.com

Volume 7, Issue 5, May 2020

Time Complexity

The time complexity comparison is illustrated in Figure 5. The proposed system has minimum delay compared to the existing CE-D2D, brute force and RSA algorithm methods. Latency can be reduced because data duplication and protection have been analyzed a minimum of times. The proposed System Reliability Based Neural Collaborative Filter (RNCF) and Distributed Secure Cloud Dataset (DSCDC) algorithm was compared with the existing algorithm. These two algorithms were applied to filter deep learning based on fragmentationtechniques and duplication techniques.

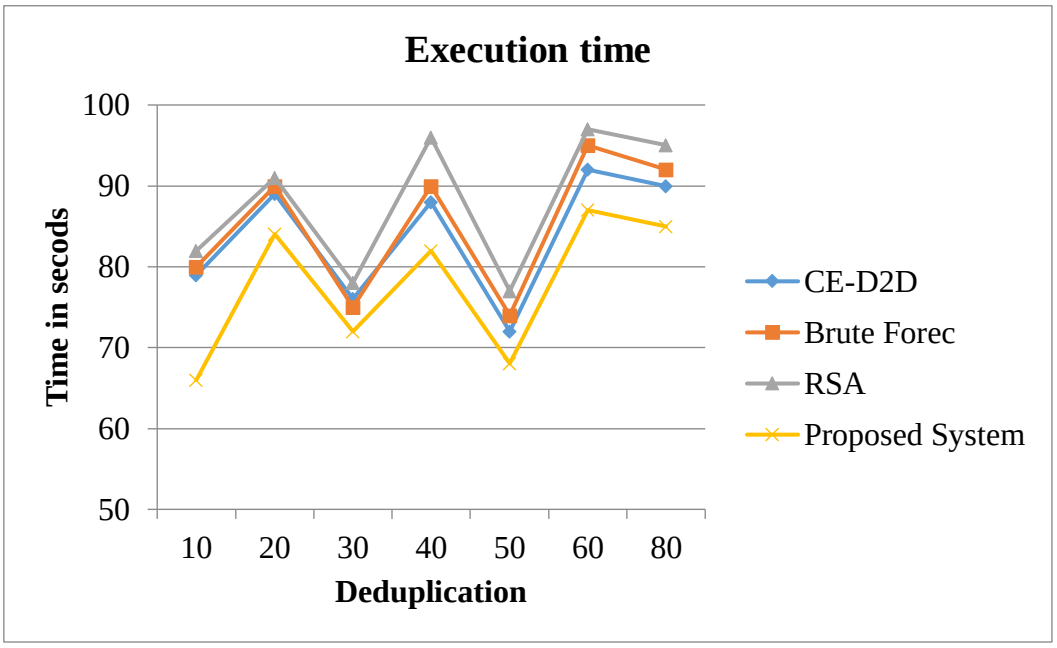


Figure 5: Comparison of Time complexity

The experimental results show that in cloud data storage, the user uploaded data was verified before uploading data. Filtering and deduplication were applied in this research. The secure data chunking and deduplication was managed in this paper. Security, reliability and efficiency were proposed in this research.

V. CONCLUSION

This research paper addresses the challenge faced in the fire level process in a secure data deduplication system. This paper focused on secure cloud storage and mainly implemented two techniques: fragmentation and data duplication. A trust-based Neural Collaborative Filter (RNCF) technique was used to block an existing cloud-stored file using granularity-, time-, and location-based blocking methods. Deduplication was applied to analyze redundant data and unique data was created specifically for the security layer. DSCDC (Distributed Secure Cloud Data Chunk) deep learning (DL) algorithm was used to authenticate the user. When a thread appeared, the algorithm block proposed in the research to ensure user safety was executed. The distributed duplication system helps to achieve the reliability, confidentiality and security of the data stored in the cloud data storage.

REFERENCES

[1] EmnaBaccour, AimanErbad, Amr Mohamed, Mohsen Guizani, and Mounir Hamdi, "CE-D2D: Collaborative and Popularity-aware Proactive Chunks Caching in Edge Networks", September 21, 2020, at 05:55:25 UTC from IEEE Xplore.
[2] Aobing Sun, Guohong Gao1, Tongkai Ji1, and Xuping Tu1, "One quantifiable security evaluation model for cloud computing platform", 2018 Sixth International Conference on Advanced Cloud and Big Data, 978-1-7281-3129-0/20/$31.00 ©2020 IEEE.

ISSN: 2395-7639

International Journal of Multidisciplinary Research in Science, Engineering, Technology & Management (IJMRSETM)

(A Monthly, Peer Reviewed Online Journal)

Visit: www.ijmrsetm.com

Volume 7, Issue 5, May 2020

[3] Wenhai Sun, Ning Zhang, Wenjing Lou, and Y. Thomas Hou, " Tapping the Potential: Secure Chunk-based Deduplication of Encrypted Data for Cloud Backup", 2018 IEEE Conference on Communications and Network Security (CNS).
[4] Xian Weiquan and Wang Houkui," The Design Research of Data Security Model Based on Public Cloud", 2013 Ninth International Conference on Computational Intelligence and Security.
[5] X. Zhang, T. Yang, C. Liu, and J. Chen, "A Scalable Two-Phase Top-Down Specialization Approach for Data Anonymization using Systems, in MapReduce on Cloud," IEEE Transactions on Parallel and Distributed, 25(2): 363-373, 2014.
[6] W. Dou, X. Zhang, J. Liu, and J. Chen, "HireSome-II: Towards Privacy-Aware Cross-Cloud Service Composition for Big Data Applications", IEEE Transactions on Parallel and Distributed Systems, 26(2): 455-466, 2015.
[7] J. Paulo and J. Pereira, "A survey and classification of storage deduplication systems", ACM CSUR, vol. 47, no. 1, p. 11, 2014.
[8] C. Yang, X. Zhang, C. Liu, J. Pei, K. Ramamohanarao, and J. Chen, "A Spatiotemporal Compression based Approach for Efficient Big Data Processing on Cloud," Journal of Computer and System Sciences (JCSS). vol. 80: 1563-1583, 2014.
[9] R. Chen, Y. Mu, G. Yang, and F. Guo, "BL-MLE: Block-Level Message-Locked Encryption for Secure Large File Deduplication", IEEE TIFS, vol. 10, no. 12, pp. 2643–2652, 2015.
[10] J. Li, C. Qin, P. Lee, and X. Zhang "Information Leakage in Encrypted Deduplication via Frequency Analysis", in Proc. of IEEE/IFIP DSN, pp. 2110–2118, 2017.
[11] Mohanasundaram, R., A. Jayanthiladevi, and G. Keerthana. "Software-Defined Cloud Infrastructure." Handbook of Research on Cloud and Fog Computing Infrastructures for Data Science. IGI Global, 2018. 108-123.
[12] D. Wu, Q. Liu, H. Wang, Q. Yang, and R. Wang, "Cache less for more: Exploiting cooperative video caching and delivery in d2d communications," IEEE Transactions on Multimedia, pp. 1788–1798, 2019.
[13] N. Zhao, X. Liu, Y. Chen, S. Zhang, Z. Li, B. Chen, and M. Alouini, "Caching d2d connections in small-cell networks," IEEE Transactions on Vehicular Technology, vol. 67, no. 12, pp. 12 326–12 338, 2018.
[14] Mxoli, Avuya, Mariana Gerber, and Nicky Mostert-Phipps. "Information security risk measures for Cloud-based Personal Health Records." Information Society (i-Society), 2014 International Conference on. IEEE, 2014.
[15] T. Condie, P. Mineiro, N. Polyzotis, and M. Weimer, "Machine learning on Big Data," Proceedings of the 29th IEEE International Conference on Data Engineering (ICDE), pp. 1242-1244, 2013.
[16] H. Zhu, Y. Cao, Q. Hu, W. Wang, T. Jiang, and Q. Zhang, "Multi-bitrate video caching for d2d-enabled cellular networks," IEEE MultiMedia, vol. 26, no. 1, pp. 10–20, 2019.
[17] Chang, Victor, Yen-Hung Kuo, and Muthu Ramachandran. "Cloud computing adoption framework: A security framework for business clouds." Future Generation Computer Systems 57 (2016): 24-41.
[18] Kaur, Hasveen, and P. S. Mann. "An Improved Hybrid Re-Encryption Scheme for Mobile Cloud Computing Environment." IJCA 162.4 (2017).
[19] Okonski, Aleksander. "Implementing Security Rules, Safeguards, and IPS tools for Private Cloud Infrastructures: GROOT: Infrastructure Security as a Service (ISaaS)." (2018).
[20] Chandni, M., et al. "Establishing trust despite attacks in cloud computing: A survey." Wireless Communications, Signal Processing and Networking (WiSPNET), 2017 International Conference on. IEEE, 2017.
[21] K. Bilal, E. Baccour, A. Erbad, A. Mohamed, and M. Guizani, "Collaborative joint caching and transcoding in mobile edge networks," Journal of Network and Computer Applications, 2019.
[22] E. Baccour, A. Erbad, K. Bilal, A. Mohamed, and M. Guizani, "Pccp: Proactive video chunks caching and processing in edge networks," Future Generation Computer Systems, vol. 105, pp. 44 – 60, 2020.
[23] Z. Qu, B. Ye, B. Tang, S. Guo, S. Lu, and W. Zhuang, "Cooperative caching for multiple bitrate videos in small cell edges," IEEE Transactions on Mobile Computing, pp. 1–1, 2019.
[24] T. Wang, Y. Sun, L. Song, and Z. Han, "Social data offloading in d2d-enhanced cellular networks by network formation games," IEEE Transactions on Wireless Communications, pp. 7004–7015, 2015.
[25] F. d. S. Moraes, K. V. Cardoso, and V. C. M. Borges, "Improving video content access with proactive d2d caching and online social networking," in IEEE ISCC, 2017, pp. 1043–1048.