Turkish Journal of Computer and Mathematics Education Vol.12 No.9 (2021),1547-1556

Research Article

Efficient and Reliable Secure Cloud Storage Schema of Block chain for Data De-duplication in Cloud

K V Panduranga Rao[1], Dr.V Krishna Reddy[2]

[1]Research Scholar CSE Dept, KL Deemed to be University
[2]Professor CSE Dept, KL Deemed to be University
pandukv@yahoo.com[1], vkrishnareddy@kluniversity.in[2]

Article History: Received: 10 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 20 April 2021

Abstract: Minimizing Storage space and bandwidth is a challenging task in **c**loud computing, duplication is the main key term in cloud computing because of large and multiples of digital data stored in single cloud server. Because of rapid usage and expansion of high amount of data, data de-duplication is a challenging task for removing replica related data which has been stored in cloud to reduce storage and space of bandwidth. Secure de-duplication is also a challenging task in cloud storage environment. Convergent encryption based secure de-duplication approach is used traditionally to explore and identify duplicates from multiple users data with de-duplication sharing and uploading data in cloud environment. Because of large volume data, conventional approaches have not meet practical implementation of de-duplication requirements in cloud storage. So we aim that block chain based cloud storage with data de-duplication is the main problem in distributed file storage environment. For predicting dynamic file de-duplication in file sharing of distributed environment, propose a Novel Block chain based Secure De- duplication Authentication scheme (NBSDAS) with high reliability and confidentiality in which the files are distributed to multiple servers and the information of files is recorded on the time-stamped block chain whose central authorities are replaced to automatically decentralize cloud contracts. This proposed approach combined with genetic algorithm to predict and de-duplication with respect to processing dynamic files individually in cloud computing. Security analysis demonstrates that our de-duplication scheme can achieve the proposed security goals while it has limited overhead proved by simulation experiments with comparison to traditional approaches.

Key Words: Cloud computing, convergent encryption, block chain, dynamic file, and genetic algorithm, secure resource provisioning, data with de-duplication de-duplication.

1. Introduction

Present days, cloud computing is an emerging asset pool which describes clients requests in real time world. It is the one which reform data with de-duplication and helps different users and real time organizations in reducing user's requests like storage, memory, processing of CPU, load management and giving an opportunity to access those requests via internet [1-4]. Cloud computing and competence provisions furnish regulars and schemes diverse facility to hoard and contract with their information in interloper server configurations. It is the do something of operating an association of distant source customers assisted on the Internet to stock up, administer & appraise info, rather than a nearby hand or an individual computer [6]. A cloud organization model speaks to a particular kind of cloud climate, fundamentally recognized by control, extent, and admission. There are 3 normal cloud sending techniques:[7] a. Public Cloud, b. Private Cloud, c. Mixture Cloud. Cloud processing is a huge scope conveyed registering worldview in which a lake of figuring belongings is accessible to clients by means of the Internet. Figuring belongings, e.g, handling power, accumulating, training, and organization data with de-duplication broadcast, are oral to cloud punters as the open public efficacy supervisions. Software as A Service (SaaS) is a process & support model mostly registered in the Cloud computing worldview [2] in storage and processing data with de-duplication in cloud. For efficient data with de-duplication storage of cloud, adaptability, flexibility in accessing data with de-duplication from cloud servers in distributed environment. Traditional frameworks store a data with de-duplication securely with repetitive in different storage systems, duplication of one document is repeated with numerous imitations, in this imitation of data with de-duplication, just checked name of the record not checked information. Data with de-duplication procedures were presented conventionally in extending productivity in storage of data with de-duplication in secure format. Data with de-duplication is the scalable approach to handle reduction of data with de-duplication storage i.e. keep one copy of document instead of keeping multiple copies with similar text. Presented copy maintain block level data with de-duplication with respect to fixed size or variation in size, present cloud services maintain user's perspective data with de-duplication to minimize or reduce computational cost.Convergent encryption is the one of the approach used for data with de-duplication with confidential storage data with de-duplication. This technique encrypts and decrypts data with de-duplication based on convergent key which is evaluated by computational hash based cryptographic values are stored for each copy of content. After generating convergent keys for encryption of content and share that cipher text to cloud server, store data with de-duplication with multiple document using single convergent key with similar

cipher text. Cipher text should be decrypted with respect to outsourced data with de-duplication owners with associated convergent keys of each document in cloud server. Another approach called baseline cryptographic approach worked based on access control policies with attribute based encryption, this approach mainly have two critical issues in identification of encrypted user's duplicated data with de-duplication. First one is, it is insufficient when multiple convergent keys are generated for multiple number users increased in outsourced cloud data with de-duplication. Second one is unreliable i.e. generating convergent keys in attribute based encryption doesn't protect data with de-duplication of users and theirs masters key, if master key lost then user lost everything.

This procedure motivates us to implement efficient approach to enable and manage reliable and secure data with de-duplication. Based on existing secure de-duplication approaches, current cloud storage system gives an efficient approach i.e. a Novel Block chain based Secure De- duplication Authentication scheme (NBSDAS) with high reliability and confidentiality in which the files are distributed to multiple servers and the information of files is recorded on the time-stamped block chain whose central authorities are replaced to automatically decentralize cloud contracts. This approach used block chain model for predicting data with de-duplication in dynamic files in storage system. This approach also performs analysis of dynamic file assignment process and evaluates efficient storage with blocks of different hash codes are generated in chain model for secure storage layout in selection of dynamic file data with de-duplication. Simulation results of proposed novel approach is evaluated and validated which can be generated from various dynamic files with respect to memory and processing of CPU loadings. Compared to traditional approaches, is has prominent performance in secure cloud storage systems.

2. Basic Preliminaries

This section describes the basic preliminaries (block chain procedure to arrange file encryption content in cloud storage) used in implemented procedure.

Block chain model

In cloud computing cloud service, different users consists consistent resources with respect to communication, storage which builds user's block chain. Data stored in IoT terminals in the form of encryption and all the users consist limited computation cost; it determines the performance in terms of computational processing cost. Computational processing cost aware communication is used in proposed approach to store data in encrypted format using block chain technology, in the process of data transmission; each and every node continues computational processing cost aware data transfer to reduce the probability of reduction of computational processing cost at every user. Based on literature relates to block chain with reduction of computational processing cost at each user uploaded files is as follows:

$$Are(j) = K(i)RE(j) / \sum_{n \in K(i)} RE(n).K(I) \quad \ldots\ldots\ldots (1)$$

Based on above representation of data in Are format, multiple users share data with sufficient could be used in encryption to recover loosed data; procedure should be stored in figure 1.

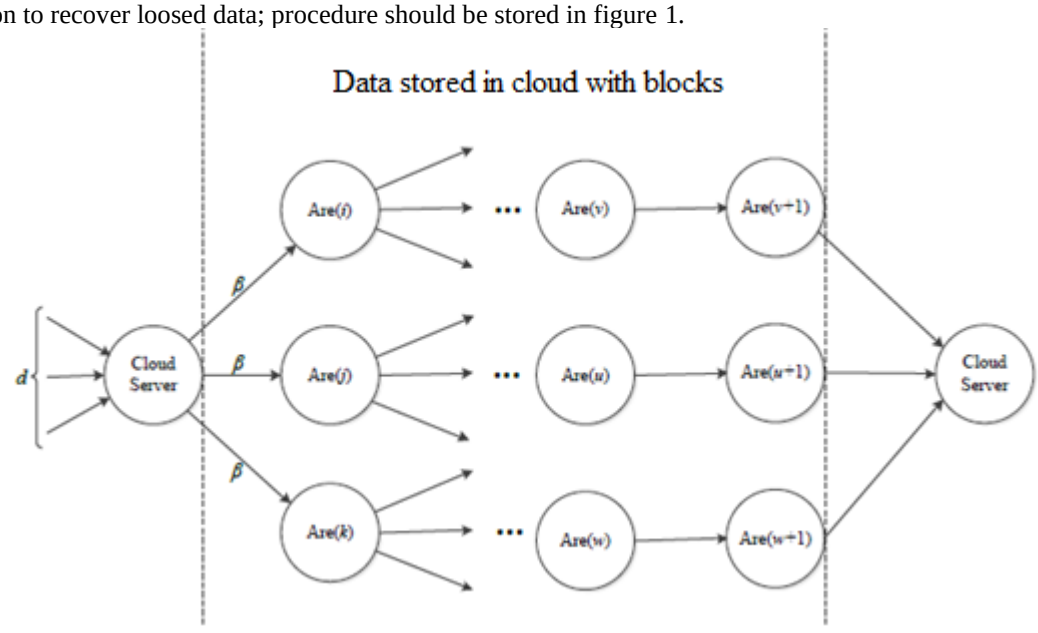

Figure 1 User's block chain based data storage

As shown in figure 4, total users equipments with respect to computational processing cost maintenance at each user to collaborative encrypted and store lose data. Data file stored in many blocks i.e. N and distributed into n local blocks, divided into n parts represented in matrix as

$$N_1 = \begin{bmatrix} x_1^1 & x_2^1 & x_{n-1}^1 & x_k^1 \\ x_1^2 & x_2^2 & x_{n-1}^2 & x_n^2 \\ L & L & L & L \\ x_1^n & x_2^n & x_{n-1}^n & x_n^n \end{bmatrix} \text{ ---------- (2)}$$

Increase the reliability of data storage cloud computing system, source data file encrypted with same code block with redundant features and represented in matrix as follows

$$N_2 = \begin{bmatrix} x_{n+1}^1 & x_{n+2}^1 & L & x_k^1 \\ x_{n+1}^2 & x_{n+2}^2 & L & x_n^2 \\ L & L & L & L \\ x_{n+1}^n & x_{n+2}^n & L & x_k^n \end{bmatrix} \text{ ---------- (3)}$$

Equations 9,10 describes N_1, N_2 represents complete data storage with completer blocks, whenever some users have been damaged then user download total data and transmitted into all the users in cloud computing based distributed environment.

3. System Design & Implementation

This section describes the design and implementation procedure of proposed approach i.e. Novel Block chain based Secure De- duplication Authentication scheme (NBSDAS). In order to make cloud storage response time to be fast and reliable, block chain based system to be used for dynamic and secure file storage. In block chain cloud system, dynamic files are stored in clocks with associated connection of different files. Basic design flow for identifying data de-duplication in cloud is shown in figure 2.

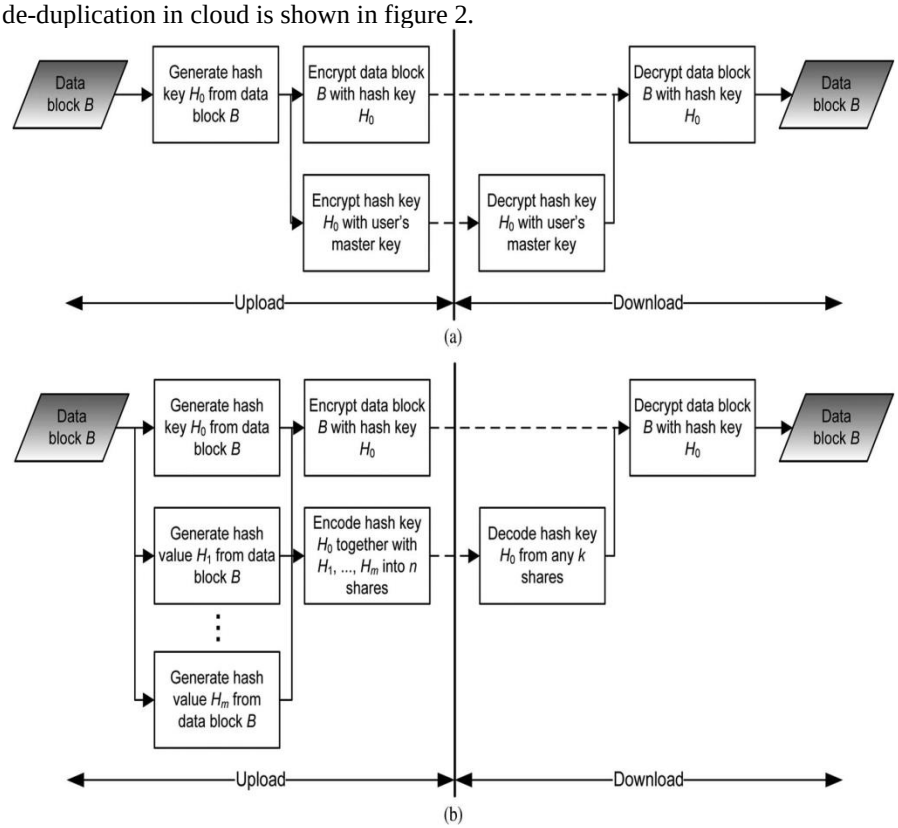

Figure 2. Step by step flow process of proposed approach.

As shown in figure 2, it describes basic flow procedure of Novel Block chain based Secure De- duplication Authentication scheme (NBSDAS), this figure also describe the ordinary file transfer and modules relates to data

de-duplication for simplification of secure data. To configure usage of multi-core processor with contemporary processors which are able to identify file processing modulations with respect to parallel running of core file operations.

To transfer a document F as the above figure shows, the information proprietor initially encodes the record with united encryption conspire.

Subsequent to utilizing K = KeyGenCE(F) produce C = EncryptCE(F, K), the information proprietor registers the record label Tag(C) and plays out a neighborhood copy check with the downloaded blockchain label data Tag(C) which is refreshed continuously.

In the event that the record has been put away, the information proprietor figures and sends Tag(C, AddrCSPi) to BSC by means of a safe channel to approve members with Register.

At that point, a TSC is endorsed by both the information proprietor and CSP. CSP distribute the TSC on blockchain and a pointer for the offer {Tag(C), Tag(ci), ci} put away at worker AddrCSPi is offered to payer after the exchange made by TSC is surrendered in the blockchain network.

In the event that no copy is discovered, the information proprietor plays out the accompanying strategy. First and foremost, the mysterious sharing calculation over C is executed and yield ci = Share(C), where ci is the I-th shard of C. And afterward, the information proprietor runs the label age calculation to get Tag(ci) and Tag(C, AddrCSPi) for every worker with AddrCSPi. From that point onward, the information proprietor transfers the arrangement of qualities {Tag(C), Tag(C, AddrCSPi)} and the marked TSC content with pre-agreed cost to the I-th CSP. At last, CSP stores these qualities at that point returns a pointer to the information proprietor while sign and broadcast the TSC to the blockchain network.

We pick 4 KB as the default information block size. A bigger information block size (e.g., 8 KB rather than 4 KB) brings about better encoding/disentangling execution because of fewer chunks being overseen, however has less capacity decrease offered by deduplication. For every information block, a hash key of size 32 bytes is created utilizing the hash work SHA-256, which has a place with the group of SHA-2 that is currently suggested by the US National Institute of Guidelines and Technology (NIST). Also, we receive the symmetric-key encryption calculation AES-256 in Cipher-Block Chaining (CBC) mode as the default encryption calculation

4. Experimental Evaluation

We evaluate the encoding and decoding performance of proposed approach on generating and recovering key shares, respectively. All our experiments were performed on an Intel Xeon E5530 (2.40 GHz) server with Windows OS. We first evaluate several basic modules that appear in proposed approach i.e. with traditional approach:

. Average time for generating a 32-byte hash from a 4 KB data block: 25.196 usec;

. Average time for encrypting a 4 KB data block with its 32-byte hash: 23.518 usec;

. Average time for decrypting a 4 KB data block with its 32-byte hash: 22.683 usec

Table 1. Different time values relates to encryption

Computational Encryption Time				
Input Files	Novel Block chain based Secure De-duplication Authentication scheme (NBSDAS)	Convergent Encryption	Efficient And Reliable Convergent Key Management	Secure Distributed De-Duplication System
10	2.56	3.342	4.424	4.352
20	4.42	5.301	6.415	7.401

30	5.472	6.399	7.28	9.338
40	7.445	8.444	9.392	11.345
50	sss.472	9.333	11.335	17.345

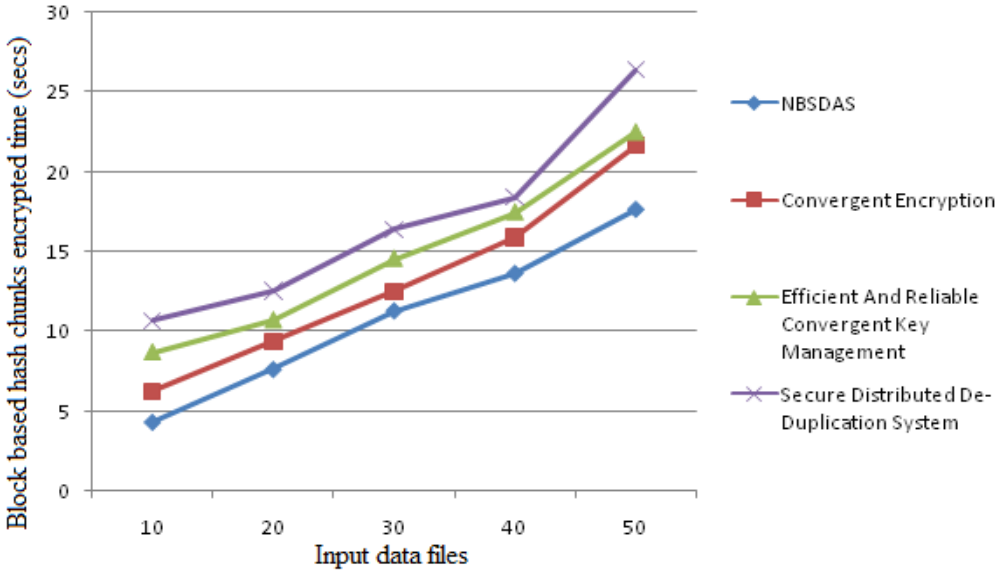

Figure 3 Performance evaluation of encryption time for dynamic file de-duplication

Table 2 Different time values relates to decryption

Computational Decryption Time				
Input Files	Novel Block chain based Secure De-duplication Authentication scheme (NBSDAS)	Convergent Encryption	Efficient And Reliable Convergent Key Management	Secure Distributed De-Duplication System
10	3.56	4.342	5.424	6.352
20	5.42	6.301	7.415	8.401
30	6.72	7.399	8.28	10.338
40	8.45	9.444	10.392	12.345
50	9.472	10.333	13.335	18.345

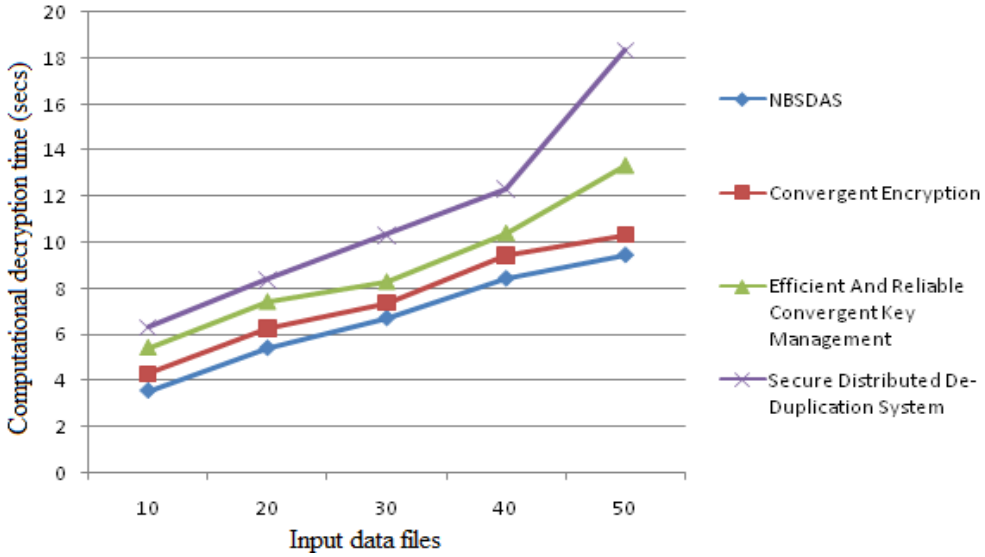

Figure 4. Performance of overall decryption time

Table 3. Overall execution values for different files processing

Computational Execution Time				
Input Files	Novel Block chain based Secure De-duplication Authentication scheme (NBSDAS)	Convergent Encryption	Efficient And Reliable Convergent Key Management	Secure Distributed De-Duplication System
10	8.56	9.342	11.42	12.35
20	9.42	12.01	14.41	16.40
30	12.72	15.99	16.28	19.81
40	14.45	19.55	23.39	22.34
50	16.72	22.44	28.33	29.34

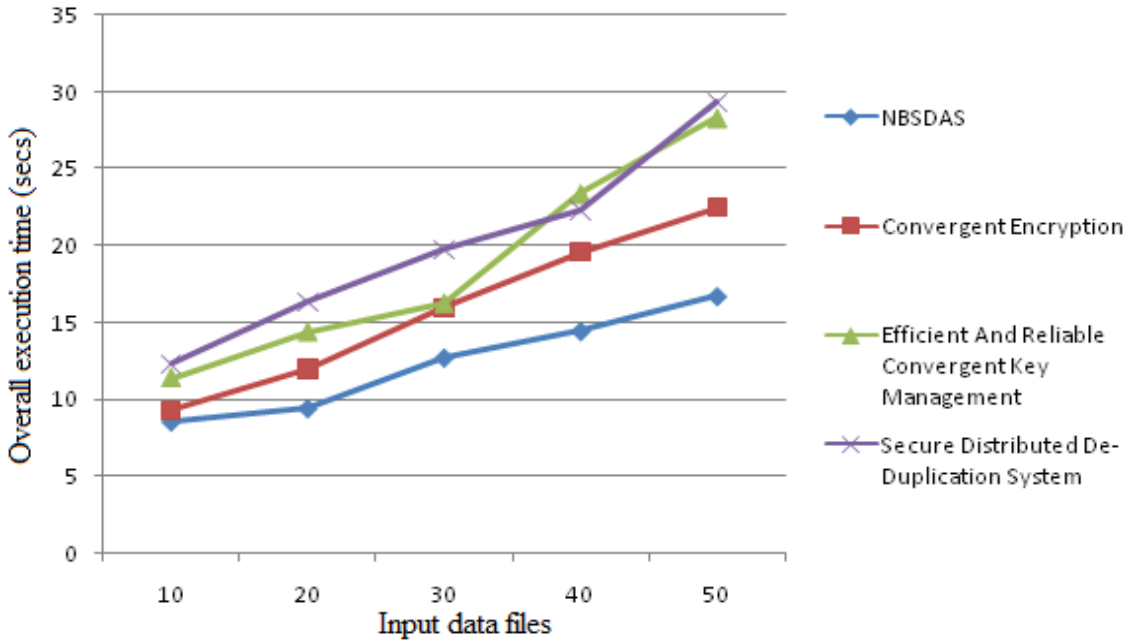

Figure 5. Performance evaluation for overall processing system

Table 4. Memory utilization values for processing de-duplication

Memory Utilization for different files				
Input Files	Novel Block chain based Secure De-duplication Authentication scheme (NBSDAS)	Convergent Encryption	Efficient And Reliable Convergent Key Management	Secure Distributed De-Duplication System
10	267	563	956	1053
20	456	1126	1256	1863
30	964	1568	1796	2169
40	1125	1863	2256	2956
50	1365	2675	3568	4256

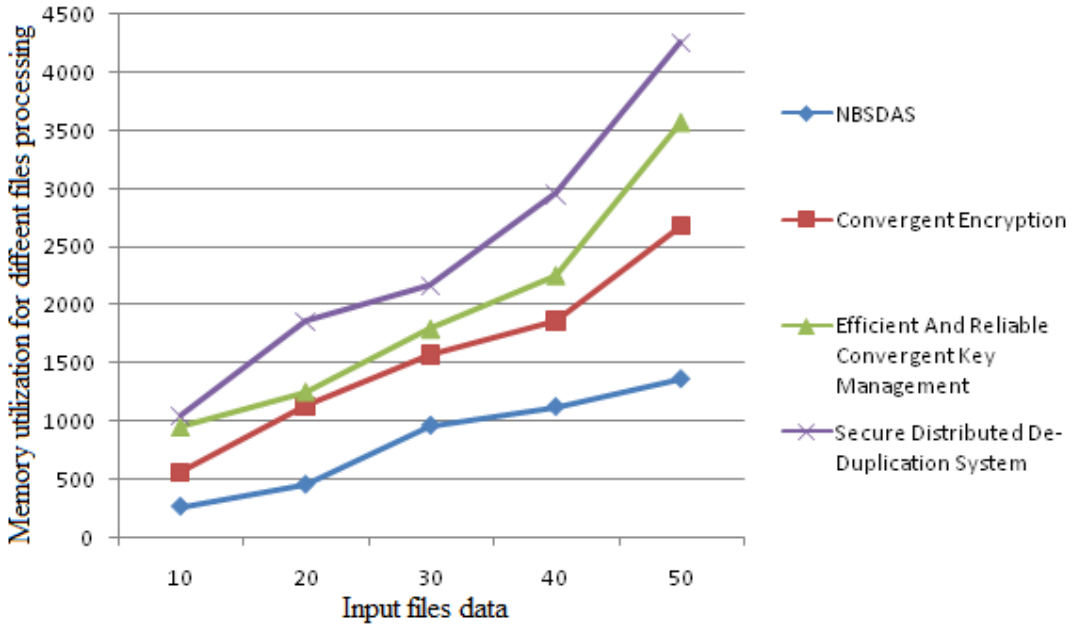

Figure 6. Performance evaluation of memory utilization in de-duplication

Table 5. Computational CPU processing cost values for de-duplication

Computation CPU processing cost				
Input Files	Novel Block chain based Secure De-duplication Authentication scheme (NBSDAS)	Convergent Encryption	Efficient And Reliable Convergent Key Management	Secure Distributed De-Duplication System
10	267	563	956	1053
20	456	1126	1256	1863
30	964	1568	1796	2169
40	1125	1863	2256	2956
50	1365	2675	3568	4256

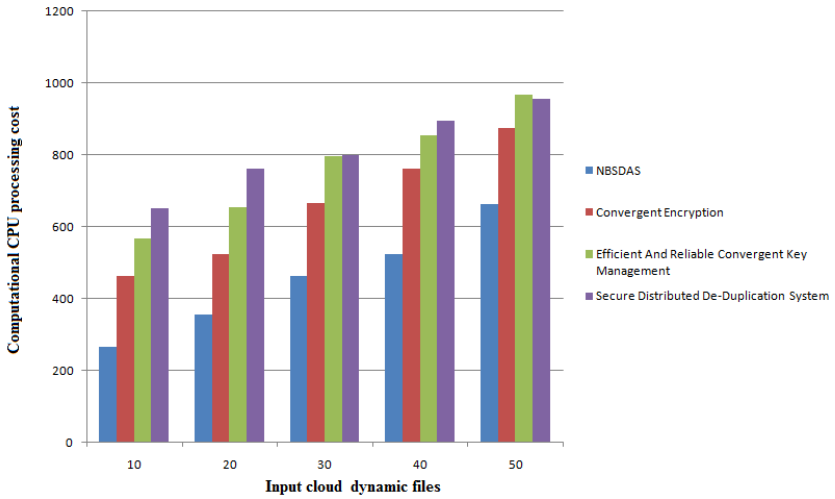

Figure 7. Performance evaluation of processing computational cost.

Based on above figures from 3-7 our proposed approach gives better results in data de-duplication with respect to time in terms of encryption and decryption, memory utilization in storage of multiple files stored and computational CPU processing cost for storage and privacy in cloud storage system.

5. Conclusion

We have proposed a block chain based distributed de-duplication scheme to improve the reliability of data storage on the promise of ensuring the confidentiality. With the aid of block chain, files are split and outsourced across multiple servers in a semi-trust decentralized storage system. Besides, auditing schemes have provided the security and integrity through cloud contract without trusted third parties which are threatened by single point of failure. Although it has incurred a little higher computation cost than baseline, security analysis has demonstrated that our proposed scheme has achieved the presented goal and provided higher security and confidentiality than the previous work.

6. References

A. Jin Li, Xiaofeng Chen, Mingqiang Li, Jingwei Li, Patrick P.C. Lee, and Wenjing Lou, "Secure Deduplication with Efficient and Reliable Convergent Key Management", IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 25, NO. 6, JUNE 2014.

B. Kamakshaiah K., Venkateswara Rao K., Subrahmanyam M. (2018) SABE: Efficient and Scalable-Filtered Access Control in Distributed Cloud Data Storage. In: Satapathy S., Bhateja V., Das S. (eds) Smart Computing and Informatics. Smart Innovation, Systems and Technologies, vol 78. Springer, Singapore. https://doi.org/10.1007/978-981-10-5547-8_4.

C. Venkatakotireddy G., Thirumala Rao B., Vurukonda N. (2018) A Review on Security Issue in Security Model of Cloud Computing Environment. In: Dash S., Naidu P., Bayindir R., Das S. (eds) Artificial Intelligence and Evolutionary Computations in Engineering Systems. Advances in Intelligent Systems and Computing, vol 668. Springer, Singapore. https://doi.org/10.1007/978-981-10-7868-2_20.

D. Chen Li and Zhenhua Liu, "A Secure Privacy-Preserving Cloud Auditing Scheme with Data Deduplication", International Journal of Network Security, Vol.21, No.2, PP.199-210, Mar. 2019 (DOI: 10.6633/IJNS.201903 21(2).03).

E. P. G., S., R. K., N., Menon, V.G. et al. A secure data deduplication system for integrated cloud-edge networks. J Cloud Comp 9, 61 (2020). https://doi.org/10.1186/s13677-020-00214-6.

F. B. Tirapathi Reddy, M. V. P. Chandra Sekhara Rao, "Filter Based Data Deduplication in Cloud Storage using Dynamic Perfect Hash Functions, DOI 10.5013/IJSSST.a.19.04.08, 2018.

G. Junbeom Hur, Dongyoung Koo, Youngjoo Shin, and Kyungtae Kang, "Secure Data Deduplication with Dynamic Ownership Management in Cloud Storage", This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/TKDE.2016.2580139, IEEE Transactions on Knowledge and Data Engineering..

H. Jin Li a, Yinghui Zhang b,c,d, Xiaofeng Chen e, Yang Xiang e, "Secure attribute-based data sharing for resource-limited users in cloud computing", computers & security 72 (2018) 1–12.

I. Naresh Vurukonda, B.Thirumala Rao, B.Tirapathi Reddy, "A Secured Cloud Data Storage with Access Privileges", International Journal of Electrical and Computer Engineering (IJECE)
J. Vol. 6, No. 5, October 2016, pp. 2338~2344.
K. Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P.C. Lee, and Weijing Lou, A Hybrid Cloud Approach for Secure Authorized Deduplication. In IEEE conference May 2015.
L. Jiali Tang 1 , Chenrong Huang 2,*, Huangxiaolie Liu 3 and Najla Al-Nabhan, "Cloud Storage Strategy of Blockchain Based on Genetic Prediction Dynamic Files", Electronics 2020, 9, 398; doi:10.3390/electronics9030398.
M. Dinh, T.T.A.; Liu, R.; Zhang, M.; Chen, G.; Ooi, B.C.;Wang, J. Untangling blockchain: A data processing view of blockchain systems. IEEE Trans. Knowl. Data Eng. 2018, 30, 1366–1385.
N. Lin, I.C.; Liao, T.C. A Survey of Blockchain Security Issues and Challenges. Int. J. Inf. Secur. 2017, 19, 653–659.
O. Pan, Z.; Yang, C.N.; Sheng, V.S.; Xiong, N.; Meng, W. Machine learning for wireless multimedia data security. Secur. Commun. Netw. 2019, 1, 1–2. [CrossRef]
P. Tapscott, D.; Tapscott, A. How blockchain will change organizations. MIT Sloan Manag. Rev. 2017, 58, 10.
Q. Bahga, A.; Madisetti, V.K. Blockchain platform for industrial internet of things. J. Syst. Softw. 2016, 9, 533–546.[CrossRef].
R. Tian, Y.; Kaleemullah, M.M.; Rodhaan, M.A.; Song, B.; Al-Dhelaan, A.; Ma, T. A privacy preserving location service for cloud-of-things system. IEEE Trans. Parallel Distrib. Syst. 2019, 123, 215–222. [CrossRef]
S. Jiang, L.; Xie, S.; Maharjan, S.; Zhang, Y. Blockchain Empowered Wireless Power Transfer for Green and Secure Internet of Things. IEEE Netw. 2019, 33, 164–171. [CrossRef].
T. Gai, K.; Wu, Y.; Zhu, L.; Xu, L.; Zhang, Y. Permissioned blockchain and edge computing empowered privacy-preserving cloud grid networks. IEEE Internet Things J. 2019, 6, 7992–8004. [CrossRef]
U. Rong, H.; Ma, T.; Cao, J.; Tian, Y.; Al-Dhelaan, A.; Al-Rodhaan, M. Deep rolling: A novel emotion prediction model for a multi-participant communication context. Inform. Sci. 2019, 488, 158–180. [CrossRef]
V. Ma, T.; Rong, H.; Hao, Y.S.; Cao, J.; Tian, Y.; Al-Rodhaan, M. A Novel Sentiment Polarity Detection Framework for Chinese. IEEE Trans. A ect. Comput. 2019, 1, 1. [CrossRef]
W. Al-Otaibi, B.; Al-Nabhan, N.; Tian, Y. Privacy-Preserving Vehicular Rogue Node Detection Scheme for Fog Computing. Sensors 2019, 19, 965. [CrossRef] [PubMed]
X. Cerf, R. The quasispecies regime for the simple genetic algorithm with roulette wheel selection. Adv. Appl.Probab. 2017, 49, 903–926. [CrossRef]
Y. 29. Cao, G.; Wang, X. Image encryption based on the combination of roulette wheel selection with linear congruence pixel transformation. Multimed. Tools Appl. 2019, 78, 1*0625–10647. [CrossRef]*.