International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-4, November 2019

An Authorized ClouDedup in Hybrid Cloud using Triple Data Encryption Standard

K Ghamya, K G Suma, V Bhargavi

Abstract-Nowadays, data deduplication has become more essential for cloud storage providers because of continuous increase in number of users and their data file size. The users are allowed to access server anytime and anywhere to upload/download their data file. Whenever data is retrieving, it leads to several problems associated to the confidentiality and privacy. For protection of data security, we proposed an efficient technique called ClouDedup which assures file deduplication. To secure the confidentiality of critical data while supporting ClouDedup checker, we proposed a triple data encryption standard (TDES) technique to encrypt the data prior to uploading the data file in cloud storage. The privilege level of user is verified with data to assure whether he is an authorized user or not. The analysis of security demonstrates that our proposed security method is safe and secure. We prove that our proposed ClouDedup method has minimal overhead compared to normal operations. The process aims to use authorized ClouDedup checker with a triple data encryption standard (TDES) technique to minimize duplication copies of data in hybrid cloud storage and conducted test experiments using our prototype.

Keywords: Hybrid cloud, ClouDedup duplication check, triple data encryption standard.

I. INTRODUCTION

Cloud storage has become so popular nowadays not in view of its effective use, however due to the availability of an oversized quantity of space for storing wherever the user will simply source their data. Most of the studies told that a portion of consumed storage on cloud is involved by duplicate data files. Challenge ahead of these days is that cloud services are that the management of that massive increasing amount of data. To make executives adaptable data management, de-duplication is presented. In deduplication, rather than putting away numerous duplicates of data which are specifically identical, it keeps just a single physical duplicate. Confidentiality and privacy are among the main considerations for public cloud situations. The user easily accomplishes data confidentiality with Triple Data Encryption standard (TDES) rather than ancient encryption and also the most widely adopted technique is provided i.e. ClouDedup for data security.

Manuscript published on November 30, 2019.
* Correspondence Author
K Ghamya*, Assistant Professor in Department Of Computer Science and Engineering from 2016 at Sree Vidyanikethan Engineering College (Autonomous), Tirupati.
Suma Kamalesh Gandhimathi, Associate Professor in Department Of Computer Science and Engineering from 2016 at Sree Vidyanikethan Engineering College (Autonomous), Tirupati.
Vemala Bhargavi, Assistant Professor in Department Of Computer Science and Engineering from 2018 At Sree Vidyanikethan Engineering College (Autonomous), Tirupati.

The basic thought about deduplication is that to store duplicate data just once. Thus, if user uploads a data file which is already in cloud, now cloud storage supplier shows deduplication is not allowed. Deduplication can reduce space for storing by up to 95% for backup applications and up to 65-68% for normal file systems [1].Users requires the protection of their data and confidentiality together with low ownership prices and adaptability which ensures based on encryption.

Unfortunately, both deduplication and encryption technologies are clashing innovations to each other. Though the purpose of Deduplication is to see copied data and to store it one time, the consequence of encryption has to build two indistinguishable data undefined subsequent to encrypt. Suppose if the user encrypted data in a normal approach, the cloud storage supplier can't have any significant bearing deduplication because two indistinguishable data were completely different after encryption. And again, if users did not encrypted data then confidentiality can't be ensured and data don't seem to be secured against attackers in cloud storage suppliers. To tackle this situation, Triple data encryption standard (TDES) approach has been proposed [2], [3], [4] and encryption key is the result of the hashing data. If we need to accomplish both deduplication and encryption at the same time, triple data encryption standard (TDES) seems to be good candidates to cope up with inherit security exposure of ClouDedup in hybrid cloud. The combined advantages of ClouDedup with triple data encryption are:

- ClouDedup deduplication guarantees file-level deduplication and secrecy of data while cooperating with triple data encryption standard (TDES) shortcomings also.
- ClouDedup deduplication conserves both privacy and data confidentiality based on extra layer of encryption even against possibly malicious cloud storage suppliers.
- ClouDedup deduplication assigns metadata manager to provide an effective key management system.
- ClouDedup deduplication mechanism works clearly with already existed cloud system. Even though, ClouDedup is completely good with APIs and also effectively coordinated with any cloud storage suppliers in our design.

The both private cloud and public cloud is called a hybrid cloud. The data which is effectively available it lives on public cloud and the most important data lives on a private cloud. Hybrid cloud provides more security to public cloud with the help of private cloud in terms of reliability, cost savings, fast deployment and extensibility [5], [6]. The difficult task of cloud storage is that to minimize oversized quantity of copied data. So, it is a procedure of removing copied data in the de-duplication approach.

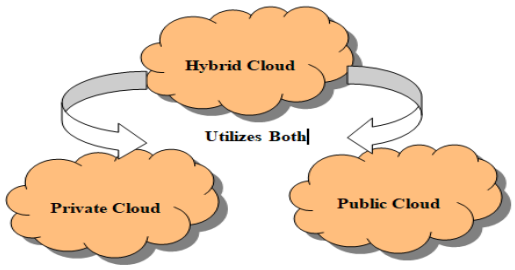

Fig. 1. Hybrid Cloud

In the existing system, the encrypted data are re-appropriated on cloud system. This encryption approach needs more time just as extra room necessity to encrypt the data if there is extensive quantity of data around then encryption approach becomes more and more difficult. Encryption approach becomes easier by applying deduplication approach in hybrid cloud. As we all of knows that the number of users are sharing oversized quantity of data through network [7]. A large network allows accessing to store and share data in cloud through network [8]. Data confidentiality and privacy of cloud could get distressed when uploading and downloading the same data number of times through the system. Hybrid cloud provides more security to public cloud by reducing redundancy in data with the help of private cloud in terms of reliability, cost savings, fast deployment and extensibility [9].

II. MOTIVATION

The most important security risk in cloud is that the data will be in plain text format whenever it stores and transfers through network. So the data will be encrypted by user and stores in cloud storage supplier. It is discovered that deduplication method can set 90% storage capacity aside, subject to applications. [10] Triple data encryption standard (TDES) has some security imperfections with relation to consistency. Hybrid cloud workloads have redundancy. A ClouDedup deduplication scheme is introduced to save the data before uploading so that storage space and disk traffic are saved.

III. LITERATURE SURVEY

In paper [5] proposed that Data de-duplication is the foremost necessary data compression procedures for removing duplicate data and also utilized in cloud to store and save bandwidth by minimizing space. To ensure the classification of private data while deduplication is upheld, the method of convergent encryption is projected to perform encryption before data is uploaded. To ensure data security, this paper concentrates more on addressing approved deduplication problem. Not at all like ancient de-duplication frameworks, differential user benefits considered in copy notwithstanding the data itself. It presents various new de-duplication builds that help approved copy check in the hybrid cloud.

In paper [6] numerous individuals currently stores a lot of information on workstations or personal PCs. These approaches have broken network frequently and also it is defenseless against equipment disappointment. Customary reinforcement planning doesn't adjust condition well and also reinforcement plans are lacking frequently. This paper portrays a method that utilizes basic information among users regarding speed backups. Convergent encryption method is utilized for scrambling individual information. It offers an interesting element that permits the quick location of data file.

In [8] presented a paper on DupLESS: The server assisted with encryption for duplicate limit. This paper, generally, features an unwavering quality and limit of the data, which has key necessities for conveyed storage. Recovery Evidence and Data Possession Test frameworks ensure the respectability of data for circulated storage. The examination of creating organizations will be overwhelming. This open-key technique in perspective on the immediate homomorphism authenticator, which enables the TPA to play out the assessment without requiring duplication of adjacent data and, in this sense, profoundly diminishes correspondence and computational overhead when contrasted and clear data looking at the methodologies.

In [11] the paper gives security tests or assaults to a substantial number of personalities on proof and mark plots explicitly which is characterized in the current writing. Fundamentally these systems clarifies how plans are going to be infer in one perspective and, on the other, permits secluded security breaks down, which helps to comprehend, improve and bring together the past work.

In [12] proposed a paper on Verify the capacity in the public cloud of secure and consistent expenses with deduplication. This paper communicates the attitude that reinforces the dependability of the data in an equipped and secure path through the de-duplication of the limit with respect to conveyed storage capacity. This tackles the issue with another arrangement in light of procedures that incorporate affirmation names dependent on polynomials and homomorphic direct authenticators. The recoverability tests (POR) and the proof of information ownership (PDP) are the techniques for the trustworthiness of data and the limit of limit with respect to disseminated capacity. Property confirmation (POW) is the technique that launches data duplicated pointlessly from the server, improving limit in a safe way. In perspective of proposing security on computational, static and strong issues of Diffie-Hellman.

IV. PROPOSED MODEL

This system includes hybrid cloud which is the combination of both public cloud and private cloud. If users stores data in public cloud then there won't be any security to user's private data and therefore, private data may be lost. So that [13] we use private cloud to provide greater security to users private data and also we use deduplication technique to remove the duplicate copies of data. In public cloud, user may upload/download their data files and therefore, private cloud provides greater security for that data files by allowing only authorized users [14]. For that, a secret key were generated by user and it is stored in the private cloud. When downloading the user's request on the private cloud to get the secret key and access that specific data files. Now let's notice the model of our system. There are three major parts/modules in our system model. They are: 1.User 2.Public cloud and 3.Private cloud.

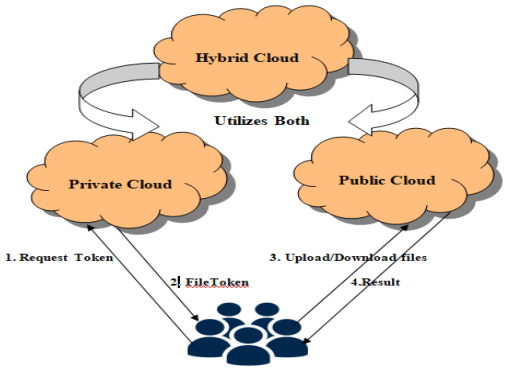

Fig. 2. Deduplication System Model

First of all, the user should initially encrypt data file with triple data encryption key before uploading in public cloud and afterwards user also creates one additional secret key for that data file and sends this secret key to private cloud for security purpose. Here, we propose a clouDedup scheme in public cloud to remove duplicate copies of data files and also it reduces bandwidth to a minimum level. In this case, the less storage space is required in public cloud to store data files. In public cloud, anyone i.e., unauthorized persons may also stores data in such a way that they can conclude that security is not provided in the public cloud. For this reason, user utilizes private cloud rather than [15] the public cloud to give greater security. And after uploading the file in public cloud, the user creates the secret key and stores it in the private cloud. At the point, when the user needs to download his/her data file then he/she sends the request message to the public cloud. Now, the public cloud gives the data files list which are uploaded by number of users in public cloud due to there is no security in the public cloud. Suppose if user chooses any one of the data file from list [16] then the private cloud communicates something specific, for example, to enter the secret key. Now the user enters the secret key he/she created at the time of file upload in public cloud. If the entered key is correct then private cloud allows the user to download the data file successfully otherwise not. Then, the user downloads and decodes the data file by using same triple data encryption key which is created when encrypting the data file. In this manner the user can utilize the model.

V. HYBRID CLOUD FOR CLOUDEDUP WITH TRIPLE DATA ENCRYPTION STANDARD (TDES)

At a high level point, enterprise network is a setting of our interest which consists of a bunch of related users who will utilizes the cloud storage suppliers and also stores the data with deduplication technique. In this configuration, deduplication can be utilized every now and again in these configuration disasters recovery and data backup applications, whereas reducing space for storing. We defined three modules in our system, they are, users, public cloud and private cloud. The cloud storage supplier performs deduplication i.e. ClouDedup technique by checking if the data in two files are identical and it stores just one of them.

The triple data encryption standard (TDES) was added to a segment put between the cloud storage suppliers and therefore the users like a portal or a local server. This segment will be in charge of encrypting, decrypting the data to and from users. To enable the cloud supplier to detect duplicate copies of data, encryption and decryption are performed with secret keys. This set of secret keys is safely put away by the segment and won't be shared to anybody under any conditions! As we are able to see, additional layer of encryption is enough to stay deduplication possible and stop the cloud supplier from acting any of the preceding attacks! In truth, the cloud supplier will never approach these secret keys. The best answer would be to create users to store their keys, we presented another part i.e., metadata manager. It acts as third party to help users with storing their encryption secret keys in private cloud and perform deduplication on encrypted data file.

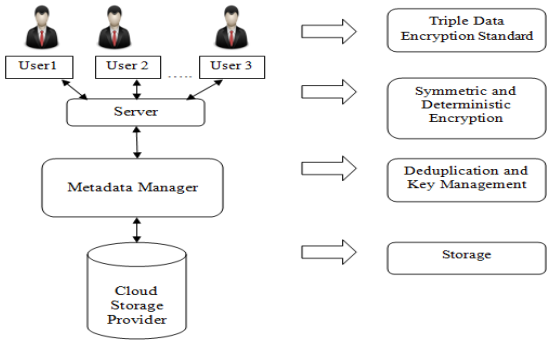

Fig. 3. High-level View of ClouDedup

Putting it all at once, our system is organized as follows:

- A user has to encrypt his/her data files before uploading to cloud with triple data encryption standard (TDES) and send to the local server encrypted data file together with their related encrypted keys.
- A local server that additionally encrypts data files and keys with a collection of distinctive and secret keys.
- A metadata supervisor/manager that updates the metadata to remake the structure of each data file and stores encrypted keys in private cloud and performs deduplication on encrypted files.
- A storage level for storing individual blocks, which can be viewed as files / small objects. Because our system is totally free of storage capacity, we can implement the storage level with any storage supplier.

VI. SYSTEM ARCHITECTURE

Each user file is related with some file tokens, which mean the predefined tag. A user measures and sends copy to check tokens to the public cloud for confirmation of approved copies. Users approach the servers in the private cloud, a reliable third party that will help you perform deduplication encryption by creating file tokens for requesting users. [17] Users also have encryption keys per user and accreditations. Therefore, a few new protection security methods are outlined against distribution attacks chosen for unpredictable messages. In other words, the definition of adjusted security ensures that the encryptions of two unpredictable messages cannot be distinguished. Therefore, the safety of data in our initial development might be ensured with this notion of security.

We have a tendency to discuss about the secrecy of data in our improved development.

In this paper, we consider only data file deduplication for ease. We refer to a copy of data in a complete file and file-level deduplication will eliminate the storage of any redundant file. [18] In reality, blocks deduplication can easily be deduced from data file deduplication. Specifically, before uploading a data file, a user initially performs duplicate verification at the data file level. In the process that the file is a duplicate, every one of its blocks should likewise be duplicated; or else, the user additionally checks for duplicates at the block level and also recognizes the distinctive blocks which will be loaded. [19] Every copy of file or a block is related to a token for duplicate verification.

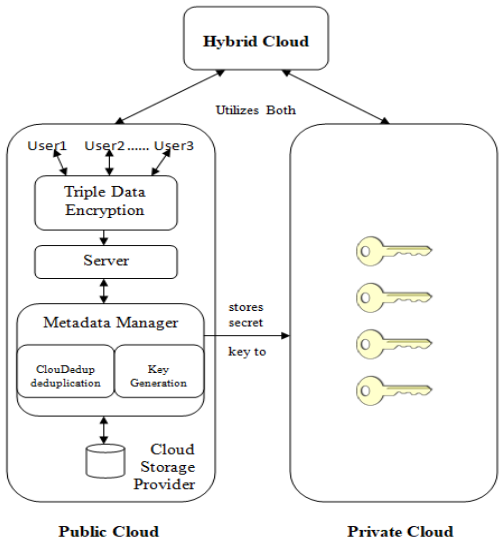

Fig. 4. Architecture of Secure Authorized Deduplication in Hybrid Cloud

Cloud Storage Supplier: This is an entity which provides data storage and outsourcing service in a public cloud on behalf of the users. Cloud storage supplier eliminates redundant data storage through deduplication and retains only distinctive data to reduce storage cost.

Users: A user is an entity that needs to re-appropriate data storage to the cloud storage supplier and in this way get to the data. In an archiving system that supports deduplication, the user loads one of a kind data, yet does not load duplicate data to spare the load bandwidth, which might be claimed by similar user or by various users. Every file is ensured with the standard triple encrypted data key.

Private Cloud: Compared to the conventional deduplication design in cloud environment, this can be a brand new entity presented to facilitate the safe use of the cloud service by the user. In particular, since the computer resources in the user / data owner are limited and the public cloud is not completely reliable in practice, the private cloud can provide the user / data owner with an execution environment and an infrastructure that works like an interface between the user and the public. Cloud with the help of the metadata manager, user secret keys is managed by the private cloud which responds to user token requests from files. The interface presented by the private cloud enables the user to send data files and inquiries to be put away and determined safely, individually. [20] Note this is another design for deduplication in cloud environment, which comprises of a double cloud (i.e., both private and public cloud). In fact, this hybrid cloud configuration has pulled in increasingly more consideration.

VII. SIMULATION RESULTS

The simulation is performed with the assistance of three steps, the primary step is to visualize whether or not the person is permitted or not, the data is reliable or not. Also second step is reserved to stay it public or not, and the last step is to take out the duplicate and build the data distinctive. Moreover, data with duplicate values is verified and keep within the cloud and this data are going to be encrypted with the assistance of standard key values of triple data encryption. Furthermore, the size of the data also will be compressed and keep as blocks.

VIII. CONCLUSION

Various new deduplication approaches that allow the authorized duplication of hybrid cloud architecture, in which the private cloud server generates keys duplicating token files. The security investigation demonstrates that our plans are safe as far as inner and outer attacks determined in the proposed security system. As a proof, we actualized a model of our proposed ClouDedup deduplication approved verification theme and conducted experiments to check our model. We have demonstrated that our approved ClouDedup duplicate check scheme brings about insignificant overhead contrasted with triple data encryption and network transfer. Hybrid cloud supplies bigger flexibility for organizations by giving choices regarding management and security. Hybrid clouds are typically enforced by associations that wish to require a part of their workload to public clouds, either for cloud explosion or for comes that need quicker implementation as a result of hybrid clouds vary in line with the wants of the business and implementation structure. Within the proposed system, the ClouDedup deduplication of approved data to ensure data security, together with the differential benefits of users within the duplicate verification system has been proposed. The proposed system is safe as far as inner and outer attacks laid out in the proposed security system. The proposed method approved duplication verification scheme which results in insignificant overhead contrasted with triple data encryption standard and network transfer.

REFERENCES

1. William J Bolosky and Dutch T Meyer. A study of practical deduplication. ACM Transactions on Storage (TOS), 7(4):14, 2012.
2. P Simon, William J Bolosky, Atul Adya, Marvin Theimer and John R Douceur. A Reclaiming space from duplicate files in a server less distributed file system. In Distributed Computing Systems, 2002. Proceedings. 22nd International Conference on, pages 617–624. IEEE, 2002.
3. Pettitt John. Hash of plaintext as key? http://cypherpunks.venona.com/date/ 1996/02/msg02013.html.
4. Freenet. https://freenetproject.org/.
5. Xiaofeng Chen, Jin Li, Yan Kit Li, Patrick P. C. Lee, and Wenjing Lou. A Hybrid Cloud Approach for Secure Authorized Deduplication.
6. L. Zhang and P. Anderson. Fast and secure laptop backups with encrypted de-duplication. In Proc. of USENIX LISA, 2010.
7. R. Kuhn and D. Ferraiolo. Role-based access controls. In 15th NIST-NCSC National Computer Security Conf., 1992.

8. T. Ristenpart, M. Bellare and S. Keelveedhi. Dupless: Serveraided encryption for deduplicated storage. In USENIX Security Symposium, 2013.
9. S. Keelveedhi, T. Ristenpart and M. Bellare. Message locked encryption and secure deduplication. In EUROCRYPT, pages 296–312, 2013.
10. P. Shah and P. Prajapati, "Efficient cross user data deduplication in remote data storage," International Conference for Convergence for Technology-2014, Pune, 2014, pp. 1-5.
11. G. Neven, M. Bellare, and C. Namprempre. Security proofs for identity-based verification and signature schemes. J. Cryptology, 22(1):1–61, 2009.
12. A. Palacio and M. Bellare. Proofs of security against impersonation under active and concurrent attacks : Gq and schnorr identification schemes. In CRYPTO, pages 162–177, 2002.
13. B. Pinkas, D. Harnik, S. Halevi and A. Shulman-Peleg. Proofs of ownership in remote storage systems. In Y. Chen, V. Shmatikov , and G. Danezis editors, ACM Conference on Computer and Communications Security, pages 491–500. ACM, 2011.
14. W. Lou, X. Chen, M. Li, J. Li, P. Lee, and J. Li. Secure deduplication with proficient and reliable convergent encryption key management. IEEE Transactions on Parallel and Distributed Systems.2013.
15. The LIBCURL. http://curl.haxx.se/libcurl/.
16. P. Lee and C. Ng. Revdedup: A reverse deduplication storage space system optimized for reads to latest backups. In Proc. of APSYS, Apr 2013.
17. S. Dorward and S. Quinlan. Venti: a new approach to archival storage. In Proc. USENIX FAST, Jan 2002.
18. L. Kencl , A. Sorniotti, E. Androulaki, and J. Stanek. A secure data deduplication scheme for cloud storage. In Technical Report, 2013.
19. D. D. E. Long, M. W. Storer, K. Greenan, and E. L. Miller. Secure data deduplication. In Proc. of StorageSS, 2008
20. H. L. Feinstein, C. E. Youman, E. J. Coyne, and R. S. Sandhu. Role-based access control models. IEEE Computer, 29:38–47, Feb 1996.

AUTHORS PROFILE

K Ghamya received B.Tech degree in Computer Science and Engineering in 2014 from Sri Venkateswara Engineering College for Women affiliated to JNTUA University, Tirupati. She has received M.Tech. degree in Computer Science and Engineering in 2016 from Sree Vidyanikethan Engineering College (Autonomous), Tirupati. She is Gold Medalist in PG. She is currently working as assistant professor in department of Computer Science and Engineering from 2016 at Sree Vidyanikethan Engineering College (Autonomous), Tirupati. Her area of research interest is Software Engineering, Network security and Cloud Computing. She published many papers in National Conferences and International journal in area of software Engineering and Network Security. She is a member of CSTA, IACSIT and Oracle Academy.

Suma Kamalesh Gandhimathi received B.E degree in Computer Science and Engineering in 2009 from SMEC affiliated to Anna University, Chennai. She has received M.E. degree in Computer Science and Engineering in 2012 from Muthayammal Engineering College, Rasipuram affiliated to Anna University, Chennai. She is the University Rank Holder in PG. she has received Ph.D. degree in Computer Science and Engineering in 2017 from Anna University, Chennai. She is currently working as associate professor in department of Computer Science and Engineering from 2016 at Sree Vidyanikethan Engineering College (Autonomous), Tirupati. Her area of research interest is Medical Image Processing. She published many papers in Conferences, National and International journal in area of Medical Image Processing. She is a member of ACM and IEEE.

Vemala Bhargavi received B.Tech degree in Computer Science and Engineering in 2012 from Vaishnavi Institute of Technology, affiliated to JNTUA University, Tirupati. She has received M.Tech. degree in Computer Science and Engineering in 2015 from S V University, Tirupati. She is University Rank Holder in PG. She is currently working as assistant professor in department of Computer Science and Engineering from 2018 at Sree Vidyanikethan Engineering College (Autonomous), Tirupati. Her area of research interest is Natural Language Processing, Network security and Cloud Computing. She published many papers in National Conferences and International journal in area of Natural Language Processing, Cloud Computing and Network Security. She is a member of CSTA, IACSIT and Oracle Academy.