ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 10, October 2019

Data Deduplication with Encryption In Cloud

Kanika S. Gandhi[1], Devashree S. Patekar[2], Garima M. Virulkar[3], Kulshree S. Damle[4],

Prof. Ms. Shwetambari G. Pundkar[5]

U.G. Student, Department of Computer Science &Engineering, Prof. Ram Meghe Institute of Technology & Research, Amravati, India[1]

U.G. Student, Department of Computer Science &Engineering, Prof. Ram Meghe Institute of Technology & Research, Amravati, India[2]

U.G. Student, Department of Computer Science &Engineering, Prof. Ram Meghe Institute of Technology & Research, Amravati, India[3]

U.G. Student, Department of Computer Science &Engineering, Prof. Ram Meghe Institute of Technology & Research, Amravati, India[4]

Associate Professor, Department of Computer Science &Engineering, Prof. Ram Meghe Institute of Technology & Research, Amravati, India[5]

ABSTRACT: Cloud Computing is an emerging technology which has become today's trending research area due its ability to reduce expenditure associated with computing. In maintaining data, Cloud Computing plays crucial role. To avoid data redundancy on cloud, we use Deduplication technique which helps to reduce duplicate data. It is necessary to maintain privacy of records that are stored on cloud because there is risk of getting exploited. There are various encryption algorithms that provide security and privacy of data.

KEYWORDS: Cloud computing, data deduplication, encryption, decryption.

I. INTRODUCTION

The amount of digital data is growing rapidly in the world. The estimated amount of data generated in 2017 and 2018 is 30 zettabytes and 33 zettabytes respectively [17]. The internet users generate 3.8 million search requests in Google search perminute.Accordingtorecentresearchthereareapproximately682milliontweetsuploadedperday.Mostofthecloudstorage space is occupied by duplicate data and the cost for handling this duplicate data is more than that of the original data [2]. To manage data storage cost-effectively, has become one of the inviting issue to be solved for the cloud service providers. To resolve this issue, Data Deduplication is broadly applied in the cloud service providers, which remove the duplicate data to enhance storage utilization and reduce and management expenditure.

Data Deduplication is an adequate data reduction approach that not only reduces storage space, by reducing duplicate data but also minimizes the transmission of the redundant data. According to data processing unit, data deduplication can be divided into file-level data deduplication and block-level data deduplication; based on data execution object, it can be divided into server-side data deduplication, client-side data deduplication and cross-user data deduplication [8]. Server-side data deduplication is a two phase process while client-side data deduplication is a one phase process.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 10, October 2019

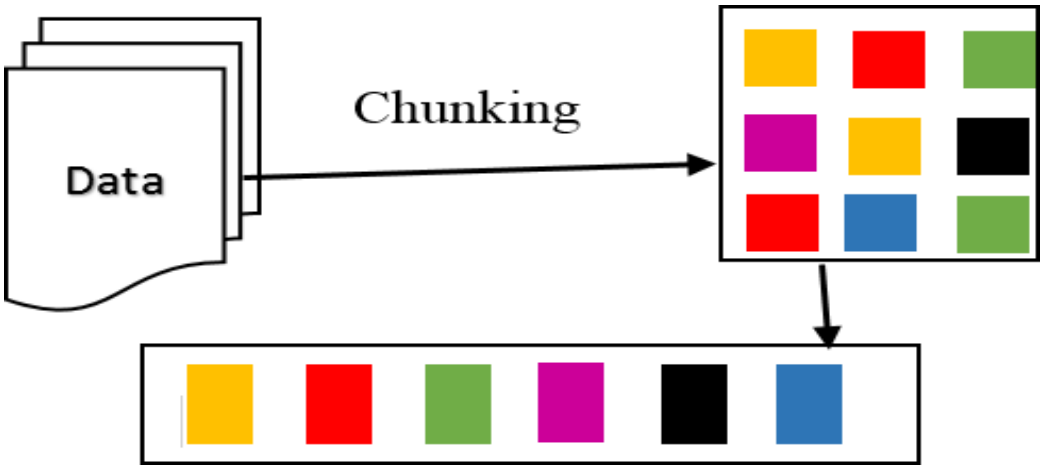

Figure: Process of Deduplication

the encrypted data defined by the sender, then user can decrypt the cipher text and retrieve the data. Here the set of attributes consist of key, which helps in encryption and decryption.

Even though data deduplication has its advantages, it also have some security challenges [10]. When the users use cloud services, they give control of their private data to the cloud service providers. This causes the risk of privacy leakage [11].

Encrypted data is more secure to transmit over insecure network and any unauthorized will not be able to read the encrypted data if he/she is not having the secret key to decrypt message. There are many encryption techniques available which have been proposed for data and network security. Encrypting data provides additional advantages other than the confidentiality and privacy of a message. These advantages assuring that messages have not been revoked during transmit and verifying the identity of the sender.

II. RELATED WORK

SHOBANA et al. [4] proposed a system that constrict the data by eliminating duplicate data, to secure the privacy of the sensitive data during deduplication. The convergent encryption method is used to encrypt the data before deploying.

Since deduplication provides considerable savings in both disk capacity and network bandwidth, D. Harnik et al. [3] examined a mechanism that provides higher privacy guarantees while slightly reducing bandwidth savings. Ng et al. [6] proposed a private data deduplication in data storage, where a client held a private data proves to a server stored a summary string of the data that he/she is the owner of that data without revealing further information to the server. Xia et al. [18] review the background of data deduplication and the differences between data deduplication and traditional data compression.

Xu et al. [7] proposed a cryptographic primitive to magnify the security of client- side deduplication in the bounded leakage setting where a certain amount of efficiently-extractable information about any file is leaked. Encrypted deduplication has been deployed in commercial cloud environments and enormously studied in the literature to concurrently achieve both data confidentiality and storage efficiency. In this paper, Li et al. [2] proposed how the deterministic nature of encrypted deduplication makes it susceptible to information leakage caused by frequency analysis.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 10, October 2019

Storer et al. [8] have developed two models for secure deduplicated storage which are authenticated and anonymous. These two models demonstrate that security can be combined with deduplication in such a way that provides a multiple range of security characteristics. In the models they presented, security is provided through the use of convergent encryption. In both the authenticated and anonymous models, a map is created for each file that describes how to reconstruct a file from blocks. To prevent information leakage, several solutions have been proposed. However, these solutions are based on a strong assumption that all individual files are independent of each other. Shin et al. [9] proposed a storage GW-based secure client-side deduplication protocol. A storage GW is a network appliance that provides access to the remote cloud server and simplifies interactions with cloud storage services, and is used in various cloud service delivery models such as public, private, and hybrid cloud computing. The proposed solution, by utilizing the storage GW as an important component in the system design, achieves greater network efficiency and architectural flexibility while also reducing the risk of information leakage.

However, the above data deduplication schemes do not take into account the key updating and user revocation. Kwon *et al*. [13] proposed a new deduplication scheme with multimedia data, which is based on randomized convergent encryption and privilege-based encryption to achieve authorized deduplication and user revocation. Hur *et al*. [14] proposed a novel data deduplication scheme for the server-side, which uses the randomized convergent encryption algorithm and ownership group key distribution technique to achieve the authorized access and support security deduplication with ownership changes dynamically. Ding *et al.* [15] proposed a secure encrypted data deduplication scheme, which exploits the homomorphic encryption algorithm to achieve security data deduplication, and supports ownership check and user revocation. However, the above schemes exploit the homomorphic encryption and proxy re-encryption with high computation cost.

In this project, the notion of authorized data deduplication was proposed by Li et al. [10] to protect the data security by including differential privileges of users in the duplicate check. In this project we perform several new deduplication constructions supporting authorized duplicate check in hybrid cloud architecture, in which the duplicate-check tokens of files are generated by the private cloud server with private keys

In order to tackle the problem that the unauthorized users can access the user information only by supplying the hash value of the _le. Halevi *et al.* [12] proposed the proof of ownership (PoW), which is an interaction protocol between client side and server side to verify the ownership of that client. In, the client and server create a Merkle Hash Tree (MHT) based on the source _le, and use a challenge-response model to verify the correct of MHT path provided by the client. Blasco et al. [5] proposed a bf-PoW scheme based on the bloom filter, which requires the certain tokens from the verified client to achieve the proof of ownership efficiently. Through the security analysis, a wide range of benchmark tests and comparison of the existing schemes, the proposed scheme greatly reduces the cost of both the client and the cloud server

Zhang et al. [11] proposed two anonymous CP-ABE schemes with fast decryption by introducing a new technique called match-then-decrypt into the decryption, in which a matching phase is added before the decryption phase. For the sake of fast decryption, special attribute secret key components are generated which allow aggregation of pairings during decryption, and hence the anonymous decryption only involves small and constant number of pairings. Extensive performance comparisons show that the novel technique match-then-decrypt can greatly improve decryption efficiency in anonymous ABE.

III. PROPOSED WORK

The algorithm automatically generates mask image without user interaction that contains only text regions to be inpainted. System architecture as shown in figure describes the working of system. When user wants to upload the data, the system checks for duplicate data. Differences among the data are referred as delta which are going to add onto the cloud. These differences are stored in the encrypted format. A sender is responsible for defining access permission

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 10, October 2019

attributes of its data for granting permission to other users while encryption of data. If a user possesses a set of attributes satisfying the access policy of the encrypted data defined by the sender, then user can decrypt the cipher text and retrieve the data. Here the set of attributes consist of key, which helps in encryption and decryption.

The aim is to propose a system in which deduplication will be implemented by storing only one chunk of data which will ultimately reduce storage space on cloud. This will also help in improving reliability and security of the documents. Security is provided as the encrypted data is more secure to transmit over insecure network and any unauthorized user won't be able to read the encrypted data if the user don't have the secret key to decrypt the data. The level of security of data depends upon the encryption key size and complexity of the encryption algorithm.

The aim is to propose a system in which deduplication will be implemented by storing only one chunk of data which will ultimately reduce storage space on cloud. This will also help in improving reliability and security of the documents.
Security is provided as theencrypteddataismoresecuretotransmitoverinsecurenetworkandanyunauthorizeduserwon'tbe able to read the encrypted data if the user don't have the secret key to decrypt the data. The level of security of data depends upon the encryption key size and complexity of the encryptionalgorithm.

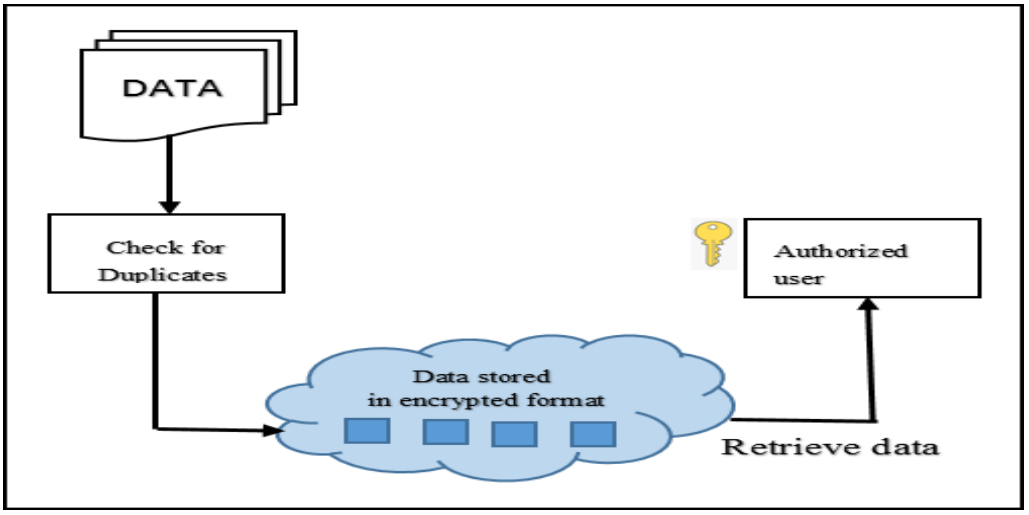

Figure: System Architecture

IV. CONCLUSION

To process the huge cloud data efficiently, the cloud service providers widely perform data deduplication to reduce the occupation of storage space and the bandwidth consumption. In order to prevent privacy data leakage, achieve authorized deduplication and satisfy dynamic privilege updating and revoking, we proposed a novel secure role re-encryption system with authorized deduplication in cloud environment. Implementation of Data Deduplication with Secured and Authorized Encryption in Cloud, this cloud based web application will reduce a lot of storage space as in this application Deduplication is used. Deduplication eliminates duplicate data stored on Cloud. Thus by reducing the storage usage, cost will also be reduced. This application is also secure as data is stored in encrypted form and at the time of downloading data proof of ownership is checked.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 10, October 2019

ACKNOWLEDGEMENT

It gives us immense pleasure to express my gratitude to Prof. Ms. S. G. Pundkar who provided me constructive criticism and a positive feedback during this seminar work.

We are indebted to Dr. G. R. Bamnote, Head of CSE department and other teaching staff who were always there whenever we needed any help. Without them and their cooperation completion of this project work would have been inevitable.

Our sincere thanks to the staff of Computer science and Engineering without their help, we could not have even conceived the accomplishment of this report. We would also like to thank the entire library staff and all those who directly or indirectly were the part of this work.

REFERENCES

1. S. G. Pundkar, G. R. Bamnote "Secure Sharing of Personal Records in Cloud using Encryption" Global Journal of Engineering Science and Researches May 2015.
2. J. Li, C. Qin, P. P. C. Lee, and X. Zhang, "Information leakage in encrypted deduplication via frequency analysis," in Proc. 47th Annu. IEEE/IFIP Int conf. Dependable Syst. Netw., Jun. 2017, pp. 1_12.
3. D. Harnik, B. Pinkas, and A. Shulman-Peleg, "Side channels in cloud services: Deduplication in cloud storage,"*IEEE Security Privacy*, vol. 8, no. 6, pp. 40_47, Nov./Dec. 2010.
4. R. SHOBANA, K. SHANTHA SHALINI, S. LEELAVATHY and V. SRIDEVI "De-Duplication of Data in Cloud" Int. J. Chem. Sci.: 14(4), 2016
5. J. Blasco, R. Di Pietro, A. Orfila, and A. Sorniotti, ''A tunable proof of ownership scheme for deduplication using bloom filters,'' in Proc. IEEE Conf. Commun. Netw. Secure. (CNS), Oct. 2014, pp. 481–489.
6. W. K. Ng, Y. Wen, and H. Zhu, ''Private data deduplication protocols in cloud storage,'' in Proc. 27th Annu. ACM Symp. Appl. Comput., 2012, pp. 441–446.
7. J. Xu, E.-C. Chang, and J. Zhou, ''Weak leakage-resilient client-side de-duplication of encrypted data in cloud storage,'' in Proc. 8th ACM SIGSAC Symp. Inf., Comput. Commun. Secur, 205, pp. 195–206.
8. M. W. Storer, K. Greenan, D. D. E. Long, and E. L. Miller, "Secure data deduplication," in *Proc. 4th ACM Int Workshop Storage Secur. Survivability*, 2008, pp. 1_10.
9. Y. Shin and K. Kim, "Differentially private client-side data deduplication protocol for cloud storage services," *Secur. Commun. Netw*, vol. 8, no. 12, pp. 2114_2123, 2015.
10. J. Li, Y. K. Li, X. Chen, P. P. C. Lee, and W. Lou, "A hybrid cloud approach for secure authorized deduplication,"
11. *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 5, pp. 1206_1216, May 2015.
12. Y. Zhang, X. Chen, J. Li, D. S. Wong, H. Li, and I. You, "Ensuring attribute privacy protection and fast decryption for outsourced data security in mobile cloud computing," *Inf. Sci.*, vol. 379, pp. 42_61, Feb. 2017.
13. S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg, ``Proofs of ownership in remote storage systems," in *Proc. 18th ACM SIGSAC Conf. Comput. Commun. Secur*, 2011, pp. 491_500.
14. H. Kwon, C. Hahn, D. Kim, and J. Hur, ``Secure deduplication for multimedia data with user revocation in cloud storage," *Multimedia Tools Appl.*, vol. 76, no. 4, pp. 5889_5903, 2017.
15. J. Hur, D. Koo, Y. Shin, and K. Kang, ``Secure data deduplication with dynamic ownership management in cloud storage," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 11, pp. 3113_3125, Nov. 2016.
16. W. Ding, Z. Yan, and R. H. Deng, ``Secure encrypted data deduplication with ownership proof and user revocation," in *Proc. Int. Conf. Algorithms Archit. Parallel Process.* Cham, Switzerland: Springer, 2017, pp. 297_312.
17. S. More, P. Shinde, S. Shaikh, V. Gunjal, S. Chavan, A. Kalia, V. Kolhe, "Secure Data Retrieval in Ad-Hoc Network using RC6 Algorithm" IJESC Vol. 6 issue no. 5, 2016.
18. W. Xia, H. Jiang, D. Feng, F. Douglis, P. Shilane, Y. Hua, M. Fu, Y. Zhang, and Y. Zhou, ``A comprehensive study of the past, present, and future of data deduplication," *Proc. IEEE*, vol. 104, no. 9, pp. 1681_1710, Sept 2018.
19. S. G. Pundkar, Dr. G. R. Bamnote "Access of Encrypted Personal Record in Cloud" International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 3 Issue: 1.