International Journal of Ethics in Engineering & Management Education
Website: www.ijeee.in (ISSN: 2348-4748, Volume 5, Issue 7, July 2018)

Review on Dynamic Ownership Management in Cloud Storage for Secure Data Deduplication

Prashansa Pathak[1], Amit Sinhal[2], Sanjay Mishra[3]
[1]M.Tech Scholar, Department of Information Technology, TIT Bhopal, prashansapathak94@gmail.com, India;
[2]Head of Department, Department of Information Technology, TIT Bhopal,
sinhal.amit@gmail.com, India;

Abstract – **Data Deduplication is one of important data mining techniques for eliminating duplicate copies of repeating data by it compares the data in cloud storage to reduce the amount of storage space and save bandwidth. Data de-duplication is used in cloud storage to save bandwidth and reduce the storage space by keeping only one copy of same data. But it raises problems involving data ownership and security when multiple users upload the same data to cloud storage. Since encryption preserves privacy, yet its randomization property hampers de-duplication. Hence, there is a need of secure data deduplication scheme to prevent unauthorized access and data leakage. In recent times, a number of deduplication schemes have been proposed to solve this problem.**

Keywords: **Deduplication, cloud storage, encryption, proof-of-ownership, revocation,**

I. INTRODUCTION

To reduce resource consumption, there are many cloud storage services, such as Dropbox, Wuala, Mozy, and Google Drive, employ a de-duplication technique. In these techniques the cloud server stores only a single copy of redundant data and provides links to the copy instead of storing other actual copies of that data. Many owners encrypt their data before outsourcing it to the cloud server to protect data privacy and protect the data from unauthorized access. A convergent encryption algorithm encrypts an input file with the hash value of the input file as an encryption key. The cipher text is given to the server and the user retains the encryption key. In the case of ownership revocation, multiple users have ownership of a cipher text outsourced in cloud storage. After some time, some of these users may request the cloud server to delete or modify their data, and then, the server deletes the owner-ship information of the users from the ownership list for the corresponding data. We propose a de-duplication scheme over encrypted data. The proposed scheme ensures that only authorized access to the shared data is possible, which is considered to be the most important challenge for efficient and secure cloud storage services in the environment where ownership changes dynamically. It is achieved by exploiting a group key management mechanism in each ownership group. As compared to the previous de-duplication schemes over encrypted data, the proposed scheme has the following advantages in terms of security and efficiency, and forward secrecy of de-duplication data upon any ownership change.[1]

Data deduplication is a specialized data compression technique for eliminating duplicate copies of repeating data in storage. The technique is used to improve storage utilization and can also be applied to network data transfers to reduce the number of bytes that must be sent. Instead of keeping multiple data copies with the same content, deduplication eliminates redundant data by keeping only one physical copy and referring other redundant data to that copy. Deduplication can take place at either the file level or the block level. For file level deduplication, it eliminates duplicate copies of the same file. Deduplication can also take place at the block level, which eliminates duplicate blocks of data that occur in non-identical files. Although data deduplication brings a lot of benefits, security and privacy concerns arise as users' sensitive data are susceptible to both insider and outsider attacks. Traditional encryption, while providing data confidentiality, is incompatible with data deduplication.[2]

Cloud Computing is a widespread term used in today's world. It delivers infinite space for storage, readiness, user-friendliness from anywhere, anytime to entities. Now-a-day's number of users and their data in the cloud is continuously growing with higher memory space and upload bandwidth. Data de-duplication used in cloud storage providers to resolve these overheads. Deduplication is a process of removing multiple copies of same data, to reduce the storage space and save bandwidth. But when same data outsourced by users to cloud storage some challenges are arises on data ownership and security for sensitive data.[3]

II. LITERATURE SURVEY

Junbeom Hur et al.[4] "Secure Data Deduplication with Dynamic Ownership Management in Cloud Storage", Dynamic ownership management is an important and challenging issue in secure deduplication over encrypted data in cloud storage. In this study, we proposed a novel secure data deduplication scheme to enhance a fine-grained ownership management by exploiting the characteristic of the cloud data management system. The proposed scheme features a re-encryption technique that enables dynamic updates upon any ownership changes in the cloud storage. Whenever an ownership change occurs in the ownership group of outsourced data, the data are re-encrypted with an immediately updated ownership group key, which is securely delivered only to the valid owners. Thus, the proposed scheme enhances data privacy and confidentiality in cloud storage against any users who do not have valid ownership of the data, as well as against an honest-but-curious cloud server. Tag consistency is also guaranteed, while the scheme allows full advantage to be

taken of efficient data deduplication over encrypted data. In terms of the communication cost, the proposed scheme is more efficient than the previous schemes, while in terms of the computation cost, taking additional 0:1-0:2 ms compared to the RCE scheme, which is negligible in practice. Therefore, the proposed scheme achieves more secure and fine-grained ownership management in cloud storage for secure and efficient data deduplication.

Dutch T. Meyer et al.[5] "A Study of Practical Deduplication", Authors studied file system data, metadata, and layout on nearly one thousand Windows file systems in a commercial environment. This new dataset contains metadata records of interest to file system designers; data content findings that will help create space efficiency techniques and data layout information useful in the evaluation and optimization of storage systems. Authors find that whole-file deduplication together with sparseness is a highly efficient means of lowering storage consumption, even in a backup scenario. It approaches the effectiveness of conventional deduplication at a much lower cost in performance and complexity. The environment we studied, despite being homogeneous, shows a large diversity in file system and file sizes. These challenges, the increase in un-structured files and an ever-deepening and more populated namespace pose significant challenge for future file system designs. However, at least one problem – that of file fragmentation, appears to be solved, provided that a machine has periods of inactivity in which defragmentation can be run.

Mark W. Storer et al.[6] "Secure Data Deduplication", Authors have developed two models for secure deduplicated storage: authenticated and anonymous. These two designs demonstrate that security can be combined with deduplication in a way that provides a diverse range of security characteristics. In the models authors present, security is provided through the use of convergent encryption. This technique, first introduced in the context of the Farsite system provides a deterministic way of generating an encryption key, such that two different users can encrypt data to the same ciphertext. In both the authenticated and anonymous models, a map is created for each file that describes how to reconstruct a file from chunks. This file is itself encrypted using a unique key. In the authenticated model, sharing of this key is managed through the use of asymmetric key pairs. In the anonymous model, storage is immutable, and file sharing is conducted by sharing the map key offline and creating a map reference for each authorized user.

Danny Harnik et al.[7] "Side channels in cloud services, the case of deduplication in cloud storage", In this article authors pointed out the potential risks of cross-user source based-deduplication. Authors described how such deduplication can be used as a side channel to reveal information about the contents of files of other users, and as a covert channel by which malicious software can communicate with the outside world, regardless the firewall settings of the attacked machine. Since deduplication offers substantial savings in both disk capacity and network bandwidth, authors suggested and analyzed a mechanism that provides higher privacy guarantees while slightly reducing bandwidth savings.

Can Wang et al.[8] "A Novel Encryption Scheme for Data Deduplication System", Data deduplication is an effective method to improve utilization of storage space and reduce bandwidth occupied by transmitting data in networks. But, the data, which are encrypted by the traditional encryption method that uses the entire file as a basic encryption unit, can hardly benefit from data deduplication because the difference of the encryption keys selected by different users and the avalanche effect of the encryption algorithm result in a very low deduplication ratio. Aiming at this problem, a novel encryption scheme for deduplication system (ESDS) is proposed in this paper. According to the method, the basic encryption unit is transformed from the entire file to the chunk, and a consistent method is used to generate the chunk encryption keys. Thus, the same plaintext chunk results in the same ciphertext chunk regardless by whom it is encrypted; and the contradiction between confidentiality and deduplication ratio is mitigated. The ESDS is a deterministic encryption algorithm and there exits a relationship between the chunk key and the chunk content to some extent; hence the method has a tradeoff between security and deduplication. But, security of the deduplication system does not merely depend on chunk keys. In other words, the leakage of partial chunk keys will not cause serious impact to security of the entire system, on condition that a user's private key and identity authentication password are not simultaneously leaked. Therefore, the tradeoff is necessary and feasible to make the ESDS compatible with the deduplication technology. The ESDS can effectively protect the confidentiality of the data during transmission and storage; thereby, it is applicable to the storage system using deduplication technology, which has the requirement of confidentiality.

III. DIFFERENT SCHEMES FOR DATA DEDUPLICATION

III.1. Convergent Encryption (CE)

In order to keep data privacy against inside cloud server as well as outside challengers, users may want their data encrypted. However, conventional encryption under different users' keys makes cross-user de-duplication impossible, since the cloud server would always see different ciphertexts, even if the data are the same, regardless of whether the encryption algorithm is deterministic. Douceur [5] introduces Convergent Encryption, which is the promising solution to this problem. In CE, a data owner derives an encryption key over data by using cryptographic hash function. Then computes the ciphertext using block cipher over data along with their encryption key. CE deletes data and keeps only encryption key after uploading ciphertext to the cloud storage. Since encryption is deterministic, on receipt of same file CE generates same ciphertext for it and the server does not store the file but instead updates meta-data to indicate it has an additional owner.

International Journal of Ethics in Engineering & Management Education
Website: www.ijeee.in (ISSN: 2348-4748, Volume 5, Issue 7, July 2018)

III.2. Message- Locked Encryption (MLE)

Bellare [6] introduces an idea of message-locked encryption (MLE), with its security approach to solving the problem of CE. He also proposed randomized convergent encryption (RCE) as one application of MLE which provides a technique to achieve secure deduplication. In RCE, initial uploader encrypts a message using a random encryption key and it results into a ciphertext refer as C1. This message encryption key is again encrypted along with a key encrypted key (KEK) which is derived from the message by using hash function and results into a ciphertext refer as C2. Here message tag is generated from the KEK, not from the ciphertext. So any owner can accept or reject the data and check the integrity of data by using tag information. In RCE, C2is used to distribute message encryption key and KEK is used as a group KEK which is shared among the same data holders.

III.3. Authorized De-duplication-Hybrid Cloud

Li [9] also proposes an authorized de-duplication scheme where differential privileges of users, as well as the data, are considered in the de-duplication procedure in a hybrid cloud environment. He presented several new de duplication 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. The figure shows the architecture of authorized de-duplication.

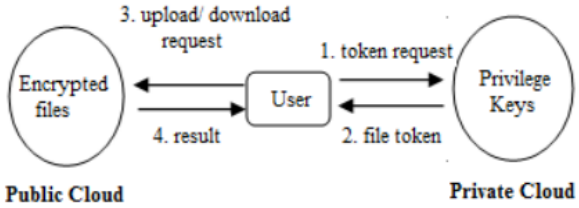

Fig.1 Authorized De-duplication model

III.4. Proxy Re-encryption Scheme (PRE)

Jin [1] introduces the scheme to address the de-duplication of encrypted data efficiently and securely with the help of ensuring the ownership of the shared file, encrypting data using keys at user's will and realizing the anonymous store through the digital credential. Proposed scheme solves the deduplication of encrypted data on the condition that no information computed from the shared data file using public algorithm used to encrypt data file. The scheme can protect clients' data by encrypting with clients' keys, and achieve secure de-duplication in encrypted data file by proxy re-encryption. The security of de-duplication against the malicious attacker is realized with MHT based verification.

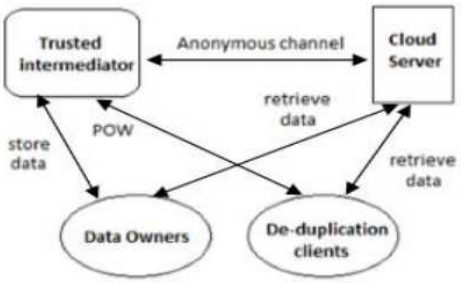

Fig.2 System model

IV. CONCLUSION

In this paper, we have reviewed different data deduplication techniques over encrypted data that are used in the cloud computing for secure data storage. Traditional encryption makes deduplication impossible because of the randomization property of encryption. Recently, several deduplication schemes are proposed to solve this issue by allowing each owner to share the same encryption key for the same data.

REFERENCES

[1] Wenxiu, Zheng Yan, and Robert H. Deng. "Secure Encrypted Data Deduplication with Ownership Proof and User Revocation." International Conference on Algorithms and Architectures for Parallel Processing. Springer, Cham, 2017.
[2] Bellare, S. Keelveedhi, T. Ristenpart, "DupLESS: Server aided encryption for deduplicated storage," Proc. USENIX Security Symposium, 2013.
[3] S. Halevi, D, Harnik, B. Pinkas, and A. Shulman-Peleg, "Proofs of ownership in remote storage systems," Proc. ACM Conference on Computer and Communications Security, pp. 491–500, 2011.
[4] Hur, Junbeom, et al. "Secure data deduplication with dynamic ownership management in cloud storage." IEEE Transactions on knowledge and data engineering 28.11 (2016): 3113-3125.
[5] Meyer, Dutch T., and William J. Bolosky. "A study of practical deduplication." ACM Transactions on Storage (TOS) 7.4 (2012): 14.
[6] Storer, Mark W., et al. "Secure data deduplication." Proceedings of the 4th ACM international workshop on Storage security and survivability.
[7] Harnik, Danny, Benny Pinkas, and Alexandra Shulman-Peleg. "Side channels in cloud services: Deduplication in cloud storage." IEEE Security & Privacy 8.6 (2010): 40-47.
[8] Wang, Can, et al. "A novel encryption scheme for data deduplication system." *Communications, Circuits and Systems (ICCCAS), 2010 International Conference on*. IEEE, 2010.
[9] Kamatchi, M. S. C., et al. "Data DEDUPLICATION Security with Dynamic Ownership Management."
[10] Y. Shin and K. Kim, "Equality predicate encryption for secure data deduplication," Proc. Conference on Information Security and Cryptology (CISC-W), pp. 64–70, 2012.
[11] Bellare, S. Keelveedhi, "Interactive message-locked encryption and secure deduplication," Proc. PKC 2015, pp. 516–538, 2015.
[12] Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P.C. Lee and Wenjing Lou , A Hybrid Cloud Approach for Secure Authorized Deduplication," IEEE transactions on Parallel and Distributed System, vol 26, No.5,May 2016.
[13] Vishalakshi N S and S.Sridevi, "Survey on Secure De-duplication with Encrypted Data for Cloud Storage," international journal of advanced science and research, Vol. 4, Issue 1, January 2017.
[14] Shweta D. Pochhi, Prof. Pradnya and V. Kasture, "Encrypted Data Storage with Deduplication Approach on Twin Cloud," vol.3 Issue-6 published at pune university in the year of June 2015.
[15] K.Kanimozhi and N. Revathi (2016) "Secure Deduplication on Hybrid Cloud Storage with Key Management," IRJET Volume: 03 Issue: 06/June 2016.