IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 1, January 2017

Survey on Dynamic Ownership Management for Secure Data De-duplication

Bhagyashri B. Nikam[1], Dr. Emmanuel M.[2]

ME Student, Department of Information Technology, PICT, Pune, India [1]

Professor, Department of Information Technology, PICT, Pune, India [2]

Abstract: 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 de-duplication schemes have been proposed to solve this problem. However, many systems suffer from security flaws because they do not reflect the dynamic changes in the ownership of outsourced data. In this paper, we review several deduplication techniques over encrypted data to achieve secure and efficient cloud storage service. Furthermore, proposed scheme uses RCE and group key management mechanism to ensure that only authorized access to the shared data is possible, which is considered to be the most important challenge for secure and efficient cloud storage service in the environment where ownership changes dynamically.

Keywords: De-duplication, cloud storage, encryption, proof-of-ownership.

I. INTRODUCTION

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. De-duplication 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. Today's cloud storage services like Dropbox and Google Drive etc. use a de-duplication scheme to save the network bandwidth and the storage cost. As data owners worried about their private data, they may encrypt their data before uploading in order to keep data privacy from illegal outside adversaries, as well as from the cloud service provider.

As concern with authorized access and security, there are many encryption schemes proposed. De-duplication scheme takes benefit of data similarity to find the same data and scale down the storage space. In contrast, encryption algorithms randomized the encrypted files to make cipher-text same from theoretically random data. Encryption of the same data by dissimilar users with different encryption keys results in different ciphertexts, which makes it hard for the cloud server to decide whether the plain data are the same and de-duplicate them. Hence, traditional encryption makes de-duplication impossible for above reasons.

The simplest implementation of traditional encryption can define as follows: Consider users A and B, encrypts the same file M under their secret keys SKA and SKB and stores corresponding cipher-text CA and CB. Then, further problems arise: First, how can the cloud server sense that the underlying file M is similar, and second is even if it can notice this, how can it allow both users to recover the stored data, based on their distinct secret keys? One simple way out is to let on each client to encrypt the file with the public key of the cloud storage server. Then, the server is capable to de-duplicate the identified data by decrypting it with its private key pair. Still, this solution grants access to the cloud storage server to get the outsourced plain data, which may break up the privacy of the data if the cloud server cannot be fully trusted.

Convergent encryption plays the vital role in data de-duplication and overcomes the drawback which discussed above. A convergent encryption algorithm works as follows: Firstly, it takes an input file and encrypts them with its hash value as an encryption key. Then, the ciphertext is given to the cloud server and user keeps the encryption key. As convergent encryption is deterministic, every time similar files encrypted into similar cipher-text irrespective of who encrypts them.Hence, the cloud server can do de-duplication over the generated ciphertext. Then all data owners can download the ciphertext and decrypt it later as they have the same encryption key for the file. But convergent encryption has security weakness concern with tag consistency and ownership revocation.This paper formalizes a scheme to solve the challenge of ownership changes dynamically in the cloud system. The proposed scheme guarantees that only authorized access to shared data is possible. It is achieved by using a group key management mechanism in each ownership group.

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 1, January 2017

The further paper is organized as follows: Section II review the literature survey for deduplication schemes. In Section III contains the conclusion.

II. LITERATURE SURVEY

In cloud computing, there have been many of the schemes, proposed for data deduplication over encrypted and unencrypted data of cloud storage. We are going to discuss about the data deduplication schemes over encrypted data and how it has been developed and improved further into Convergent Encryption (CE), Leakage-Resilient (LR) Deduplication scheme, Randomized Convergent Encryption (RCE) and Dynamic Ownership Management Scheme.

A. 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 [2] 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.

Merits: Provides promising solution over conventional encryption and preserves data privacy.

Limitations: Convergent Encryption suffers from some security issues i.e. tag consistency problem. It means that integrity and security of data has been compromised due to the lack of PoW process and dynamic ownership management.

B. Message- Locked Encryption (MLE)

Bellare [3] 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 de-duplication. In RCE, initial uploader encrypts a message using a random encryption key and it results into a ciphertext refer as C_1. 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 C_2. 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, $C_2$is used to distribute message encryption key and KEK is used as a group KEK which is shared among the same data holders.

Merits: Overcome the drawback of convergent encryption by allowing use of tag information to check data integrity.

Limitations: MLE also suffers from security issues. It does not support dynamic ownership management in data owners.

C. Leakage- Resilient (LR) De-duplication Scheme

Xu[4] proposes a leakage-resilient de-duplication scheme to solve the data integrity issue. It addressed a vital security concern in cross-user client-side de-duplication of encrypted files in the cloud storage: privacy of users' sensitive files against both outside challengers and the honestbut-curious cloud storage server in the bounded leakage model. LR also enables the use of randomly selected key to encrypt the data. Then the data encrypted key is encrypted under a KEK which is derived from the data and distributed among the data holders after the Proof-of-ownership (PoW) process. Data integrity is checked by using the data encryption key with same KEK.

Merits: Resolves data integrity problem i.e. prevents tag consistency attack.

Limitations: Secure proof of ownership (PoW) scheme in the standard model remains an open problem. Another drawback is the lack of dynamic ownership management among the data holders.

D. Ramp Secret Sharing Scheme (RSSS)

Li [5] formalizes a convergent key management scheme i.e. Dekey which is efficient and reliable for secure de-duplication. Dekey set de-duplication between convergent keys and distributes those keys across multiple key servers while preserving the semantic security of convergent keys and privacy of outsourced data.Dekey is implemented using the Ramp secret sharing scheme. Dekey uses RSSS to collect convergent keys. Its idea is to permit de-duplication in convergent keys and distribute the convergent keys over various KM-CSPs. Instead of encrypting the convergent keys on a per-user basis, Dekey builds secret shares on the original convergent keys (that are in plain) and assigns the shares over various KM-CSPs. If many users share the identical block, they can access the same corresponding convergent key. This significantly decrease the storage overhead for convergent keys. In addition, this method provides fault tolerance and allows the convergent keys to remains accessible even if any subset of KM-CSPs fails.

Merits: Provides reliable, efficient and fault tolerance convergent key mechanism for secure de-duplication.

Limitation: This scheme does not support dynamic ownership management issue in secure de-duplication.

E. Authorized De-duplication-Hybrid Cloud

Li [6] 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-

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 1, January 2017

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.

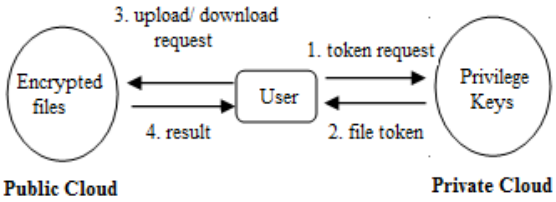

Fig 1: Authorized De-duplication model [6]

Merits: This scheme provides authorized de-duplication over hybrid cloud for users who have different privileges.

F. Proxy Re-encryption Scheme (PRE)

Jin [7] 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.

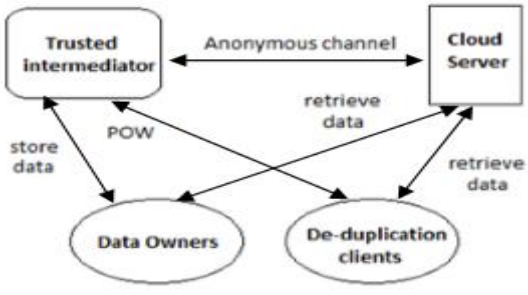

Fig 2: System model [7]

In addition, we adopt digital credential which makes that client store data file anonymously and incentive mechanisms which ensure that the scheme can execute smoothly. System model is shown in following figure.

Merits: Provides more security against malicious attacks. Hence, this technique is efficient and more secure**.**

G. Sever-aides MLE(based on CE)

Bellare [8] proposes a server-aided MLE which is secure against the brute-force attack, which was recently extended to interactive MLE to give privacy for messages that are both correlated and dependent on the public system parameters. He designs a DupLESS model that consists a CE-type base MLE scheme with the ability to obtain message-derived keys with the aid of a key server (KS) and shared among the group of clients. The patron interacts with the KS by a protocol for oblivious PRFs, guarantee that the KS can cryptographically combine in secret material guarantee that the KS can cryptographically combine in secret material.

Merits: Provides secure de-duplicated storage resisting brute-force attacks.

Limitations: This scheme does not handle the dynamic ownership management issues involved in secure de-duplication for shared outsourced data.

H. Predicate Encryption

Shin [9]formalizes a de-duplication scheme over encrypted data that uses predicate encryption.This approach allows de-duplication only of files that belong to the same user, which severely reduces the effect of de-duplication.

Merits: This scheme allows de-duplication over the files uploaded by same user so it helps to reduce the impact of removing same files.

Limitations: Scheme does not focus on deduplication across different users which providesmore storage savings and also lacking with dynamic ownership management issue.

Table I shows the comparison between various techniques of data deduplication by considering three parameters i.e. Efficiency, Tag consistency and ownership management.

Table I: Comparison of Data De-duplication Techniques

Parameters v/s Techniques	Efficiency	Tag Consistency	Ownership Management
CE	Less efficient than proposed scheme	No	No
MLE	More efficient than CE	Yes	No
LR	Efficient as it prevents data leakage	Yes	No
RSSS	Efficient and Reliable	Yes	No
Authorized de-duplication	Efficient and secure in terms of insider and outsider attacks	-	No
PRE	Efficient and secure	-	No
Server-aided MLE	Efficient and secure against the brute-force attack	Yes	No
Predicate Encryption	Less Efficient than CE	-	No
Proposed Scheme	Highly efficient and more secure than all	Yes	Yes

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 1, January 2017

III. 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. Convergent encryption has different encryption variants for secure deduplication which was formalized as MLE later in. Though, CE suffers from security flaws with regard to tag consistency and ownership revocation. The schemes MLE and LR areproposed to recover the drawback of CE but still these schemes are insecure in the setting of PoW and Dynamic Ownership Management among the data owners. Furthermore, many schemes could not achieve secure access control under dynamic environment. Hence, not much work has yet been done to address dynamic ownership management and its related security problem. Thus 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.

REFERENCES

[1] J. Li, X. Chen, X. Huang, S. Tang, Y. Xiang, M. Hassan, and A. Alelaiwi, "Secure Distributed Deduplication Systems withImproved Reliability," IEEE Transactions on Computer, Vol. 64, No. 2, pp. 3569–3579, 2015.

[2] R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M.Theimer, "Reclaiming space from duplicate files in a server less distributed file system," Proc. International Conference on Distributed Computing Systems (ICDCS), pp. 617–624, 2002.

[3] M. Bellare, S. Keelveedhi, and T. Ristenpart, "Message-locked encryption and secure deduplication," Proc. Eurocrypt 2013, LNCS 7881, pp. 296–312, 2013. Cryptology ePrint Archive, Report 2012/631, 2012.

[4] Xu, E. Chang, and J. Zhou, "Leakage-resilient client-side deduplication of encrypted data in cloud storage," ePrint, IACR, http://eprint.iacr.org/2011/538.

[5] Li, X. Chen, M. Li, J. Li, P. Lee, and W. Lou, "Secure deduplication with efficient and reliable convergent key management," IEEE Transactions on Parallel and Distributed Systems, Vol. 25, No. 6, 2014.

[6] J. Li, Y. K. Li, X. Chen, P. Lee, and W. Lou, "A hybrid cloud approach for secure authorized deduplication," IEEE Transactions on Parallel and Distributed Systems, Vol. 26, No. 5, pp. 1206–1216, 2015.

[7] X. Jin, L. Wei, M.Yu, N.Yu and J. Sun, "Anonymous deduplication of encrypted data with proof of ownership in cloud storage," Proc. IEEE Conf. Communications in China (ICCC), pp.224-229, 2013.

[8] Bellare, S. Keelveedhi, T. Ristenpart, "DupLESS: Server aided encryption for deduplicated storage," Proc. USENIX Security Symposium, 2013.

[9] 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.

[10] 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.

[11] Bellare, S. Keelveedhi, "Interactive message-locked encryption and secure deduplication," Proc. PKC 2015, pp. 516–538, 2015.

[12] C. Wang, Z. Qin, J. Peng, and J. Wang, "A novel encryption scheme for data deduplication system," Proc. International Conference on Communications, Circuits and Sytems (ICCCAS), pp. 265–269, 2010.

BIOGRAPHY

Bhagyashri B. Nikam, received BE degree in Computer Science & Engineering, from North Maharashtra University, Jalgaon in 2013. She is currently pursuing ME Degree in Department of Information Technology from Pune Institute of computer technology, Pune. Her area of interest includes Cloud Computing.

Dr. Emmanuel M, Professor of Information Technology Department, Pune Institute of Computer Technology, Pune. He has received M. Tech. degree in Computer Science& Engineering and completed Ph. D. in Computer Science& Engineering. His research interest is Big Data and Medical Image Processing.