IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Special Issue-12, May 2019 E-ISSN: 2347-2693

Secure Image Deduplication Using DICE Protocol for Reducing Storage Cost

Ketakee Dangre[1*], Amit Pampatwar[2], Raana Syeda[3]

M.Tech, Department of Computer Science & Engineering, Jhulelal Institute of Technology, Nagpur, India

Corresponding Author: ketakee.dangre16@gmail.com

Available online at: www.ijcseonline.org

***Abstract*—** Data deduplication is an important aspect for Cloud Storage Providers (CSPs) since it allows them to remove the identical data from their storage successfully. Some techniques have been proposed in the literature for this research area. Among these techniques, the Message Locked Encryption (MLE) scheme is frequently mentioned. Researchers have introduced Message locked Encryption based protocols which provide secured deduplication of data, where the data is in text form. As a result, multimedia data such as images, which are larger in size compared to text files, have not been given much attention. Applying secured data deduplication to such data files could significantly reduce the cost and space required for their storage. This helps in reducing maintenance cost as well for the storage providers. There are several other techniques for Data deduplication. Few of those techniques are as follows SPSD (Secure Perceptual Similarity Deduplication), CSPD (Client Based Secure Provable Deduplication), Image Compression.

***Keywords*—** Image Deduplication, Data Security, Cloud Storage

I. INTRODUCTION

Most of the users now a day are using cloud based services in day to day life. For Cloud storage service's deduplication has became an important technology. The term deduplication refers to techniques that eliminate duplicate copies in cloud and replace them with a pointer to the unique copy.

In this paper we compare various deduplication techniques with DICE protocol which present a secure way of deduplication scheme for near identical (NI) images with the Dual Integrity Convergent Encryption (DICE) protocol, which is a variant of the MLE (Message Locked Encryption) based scheme. In the proposed scheme the blocks that are common between two or more NI images are stored only once in the storage. As compared to other techniques DICE protocol saves large amount of memory on cloud storage as many techniques perform deduplication on entire image but in DICE protocol an image is disintegrated into blocks and the DICE protocol is applied on each and every block separately rather than on the entire image. In this we use secure block level image deduplication method that eliminates the near identical images in encrypted form, thus protecting the confidentiality of the images.

Our core idea is to divide the image into blocks and employ the DICE protocol on each block separately. Each block is encrypted using AES with a key that is obtained by hashing the image blocks.

II. RELATED WORK

Deduplication technique was already a part of research as the demand for optimum storage utilization was increasing. There are few deduplication technique researches that were carried out in past such as :

1. Message-Locked Encryption and Secure Deduplication
2. A secure cloud storage system supporting privacy-preserving fuzzy deduplication (SPSD)
3. A Client-based Secure Deduplication of Multimedia Data
4. Secure image deduplication through image compression
5. A Dual Integrity Convergent Encryption Protocol for Client Side Secure Data Deduplication

Message-Locked Encryption and Secure Deduplication

Deduplication and encryption are two opposing techniques in the sense, when the same file used by two different users is encrypted with their respective keys, the encrypted files are no longer the same, this makes deduplication quite challenging.

To overcome the issue of Convergent Encryption, MLE scheme was introduced. Convergent encryption is also acknowledged as content hash keying that creates the same ciphertext from an identical plaintext file.

MLE based scheme provide different encryption strategies such as HCE1, HCE2 and Randomized Convergent Encryption (RCE)

Convergent Encryption (CE)

In the upload protocol of the CE scheme, as shown in Figure 1 , a client calculate key $K \leftarrow H(M)$ from M and generates $C \leftarrow E(K, M)$, such that M is a plaintext message, C is a ciphertext message, K is a message-driven key, H is a cryptographic hash function, and E is a block cipher for encryption strategy. The client then uploads C to the server and stores the value of K. While receiving C, the server creates a tag $T \leftarrow H(C)$ computed using the value of C and compares T with the tags in its storage. At this stage, tag T is used as an identifier of C to check the uniqueness of C. If no tag matches T from already existing data, the server stores both C and T. Otherwise, the server will not store C; rather the server will update the meta-data to indicate that the client owns C. Later on, CE scheme creates T on server-side and determines the duplication of C. To allow the client to retrieve ciphertext C at a later time, the server returns tag T to store at the client side.

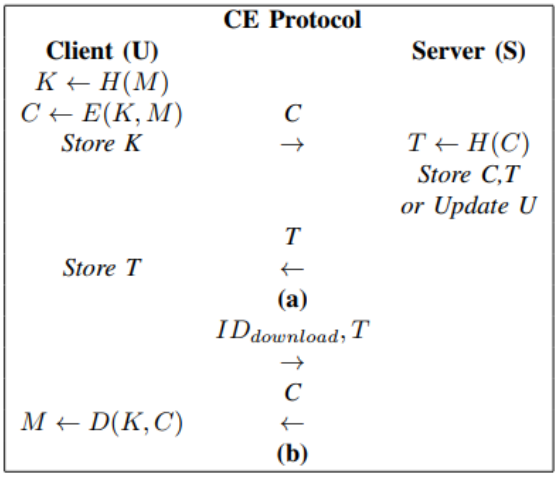

Figure 1. CE: Upload and Download

Figure 1 shows the download protocol of the CE scheme. The client can download ciphertext C by sending T, and therefore recover plaintext message $M \leftarrow D(K, C)$ by deciphering C, using its retained K, where D is a decryption scheme used to get original plain text . The CE scheme is secure against the poison attack, but due to the transmission of the duplicate file multiple times, the scheme results in a huge consumption of network resources and bandwidth requirements.

Convergent encryption has a drawback it does not work for secure fuzzy deduplication. Along with this HCE1, HCE2 and RCE techniques were used but this techniques works on text data which will not work for multimedia data like image. For implementing secure deduplication on image further techniques were introduced.

In this review paper we focus on the other technique of deduplication under group applications.

Secure Perceptual Similarity Deduplication scheme (SPSD)

It supports privacy-preserving fuzzy deduplication.

The scheme mainly consists of three aspects:

1. The perceptual hash algorithm (pHash) is introduced to generate the signatures of the images. pHash shows a good performance in measuring the similarity of perceptual similar images
2. To prevent data leakage, the images and signatures stored in the cloud service are all encrypted using the sharing group key by a symmetric cryptosystem
3. The duplicate check is completed on the encrypted pHash by calculating the Hamming distances, which determines whether the new image is to be uploaded.

The calculations of encrypted images and pHash are carried out on the user side while the duplicate check is performed on the cloud side. This scheme is based on similarity measurement which determines whether two images duplicate in the context of similarity. It does lead to storage and bandwidth savings more significantly than the traditional deduplication methods.

SPSD focuses on the group service provided by the cloud and devotes to complete deduplication safely and efficiently under this circumstance

SPSD method takes both security and efficiency into consideration while completes the fuzzy image deduplication.

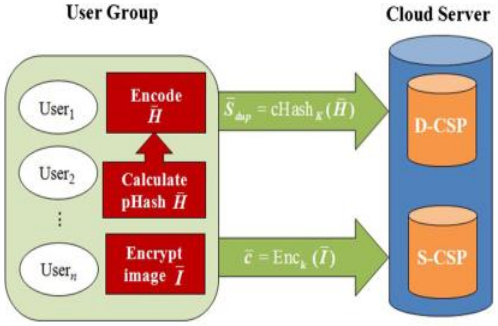

Figure 2: The framework of secure similarity deduplication method SPSD

In this model there are three entities

1. The User Group: This entity constitutes users who are authorized to outsource and share data with each other in the same group. To protect privacy, the group members agree on a secret key and use it as the group key.
2. The storage cloud service provider (S-CSP): This entity is a remote server that provides outsourcing storage services. The clients transfer data to SCSP for backup and download data from S-CSP when necessary. Since the storage data in S-CSP are encrypted with the group key, only the group members can decrypt it.

3. The deduplication cloud service provider (D-CSP) : To prevent data leakage, the pHash is encrypted before being uploaded to D-CSP. For each user group, a database of duplicate strings is established in D-CSP If a group member attempts to upload an image to S-CSP, it is first required to send the duplicate string to DCSP. Then D-CSP checks whether any similar string has already existed in the database, as it may be previously uploaded by other group member. If the check is passed, D-CSP will inform the user to upload the image; otherwise, D-CSP informs S-CSP to return the user a pointer to the perceptual similar image.

SPSD completes similarity deduplication on the encrypted data using the cryptographic hash mapping function.

Client-based Security Provable Deduplication of Multimedia Data (CSPD)

CSPD can check duplicates accurately and assess the perceptual quality of distorted images. CSPD can only perform the deduplication of identical images. It provide rigorous security analysis and extensive performance evaluation to show that the CSPD scheme meets provable security requirements and it can check duplicates accurately and store the image which has the best perceptual image quality on the server.

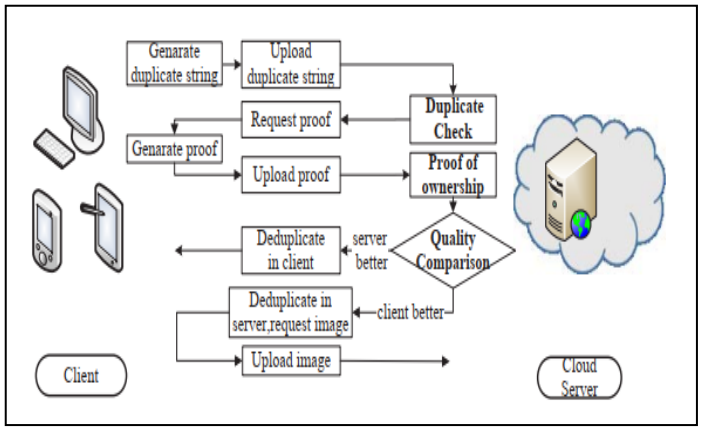

Figure3: Framework of CSPD scheme

CSPD scheme consists of three parts:
1) Duplicate Check; 2) Proof of Ownership; 3) Quality Comparison
Whereas in SPSD it doesn't provide Proof of Ownership and Quality Comparison Features. Compared with the SPSD, the CSPD can check duplicates more accurately. Furthermore, it meets provable security requirements without increasing excessive computation and time.

Secure image deduplication through image compression

Compression scheme achieves a secure deduplication of images in the cloud storage. Its pattern consists of embedding a partial encryption and a unique image hashing into the Set Partitioning In Hierarchical Trees (SPIHT) compression algorithm.

This scheme is composed of three components:
1. SPIHT compression algorithm
2. Partial encryption
3. Hashing.

The compressed image obtained from the SPIHT algorithm is partially encrypted in such a way that it is not available in plaintext form to the semi-honest CSP or any malicious user, hence ensuring the security of the data from the CSP. The image hashing technique allows a classification of the identical compressed and encrypted images based on their short signatures, in such a way that the image deduplication step is carried out efficiently.

Image Compression scheme is strong enough to identify minor changes between two images even if they are in the compressed form. This scheme is secured against the semi-honest CSP since the CSP does not have access to the compressed images, but can identify the identical images from different users only through the significant maps of these compressed images

III. METHODOLOGY

Dual Integrity Convergent Encryption (DICE) protocol

Convergent encryption is a encryption technique that creates identical ciphertext from an identical plaintext file. It contains some applications in cloud computing to eliminate all duplicate files from cloud storage, without a source needing to have access to encryption keys. In above schemes all the techniques have their own advantages over other but in DICE it works on their bandwidth and storage requirements. As above techniques MLE works on File level Deduplication It works for text file and other techniques SPSD, CSPD, Image Compression are works on entire image where as DICE divide the image into blocks and employ the DICE protocol on each block separately. Each block is encrypted using AES (Advanced Encryption Standard) with a key that is obtained by hashing the image blocks.

In DICE protocol The user first divides the image into a fixed number of blocks. Each block size could be of variable length, anywhere from 4×4, 8×8 to 16×16. After converting the image into blocks, the user runs the client portion of the DICE protocol on each block. the client computes the key Ki as $Ki \leftarrow H(Bi)$ where H is the hash function, and Bi is the i^{th} block of the image. Next, the client computes the ciphertext Ci as $Ci \leftarrow E(Ki, Bi)$ and the tag T as $Ti \leftarrow H(Ci)$, where E is the encryption strategy.

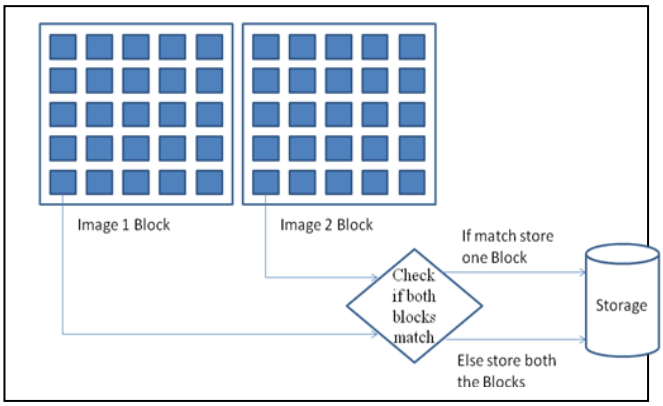

Figure 4: Block Level Deduplication of two Images

At this stage, the user sends the tag to the CSP and checks for its existence in the cloud. The CSP then runs a search in the tag store for the existence of the tags from the tag vector and sends a request for only those blocks for which no match was found. The client then sends the ciphertext of those particular blocks to the CSP, who stores them along with the user's credentials and updates its tag store by computing T'←H(Ci)

At the time of download, the user sends the tag vector and userid, and the CSP searches its tag store to find the corresponding tag and ciphertext block as Ti = Ti. If there is a match found, and then the corresponding ciphertext block is sent to the respective user, otherwise the CSP sends an acknowledgement that the image is not found.

IV. RESULTS AND DISCUSSION

1) User has to upload image in cloud based storage.
2) DICE protocol will be applied and image will be disintegrated in to blocks.
3) Each block will be verified if it is similar to any other block/reference.
4) If yes then the reference will be stored or if the block is new then the block will be stored.

V CONCLUSION AND FUTURE SCOPE

In this paper we provided a method to perform secure image deduplication at the block level based on the DICE protocol. We found that the greater the similarity of the images, the smaller the number of blocks stored at the cloud. However, the constraint here was that the images were nearly identical with small variations among them. In the future we would like to address this issue on a broader spectrum where we add more image operations like scaling, rotation, cropping, multiple viewpoints, lighting conditions and compression with different file formats, and tests them at the cloud.

REFERENCES

[1]. Ashish Agarwala, Priyanka Singh, Pradeep K. Atrey, "Client Side Secure Image Deduplication Using DICE Protocol" in 2018 IEEE Conference on Multimedia Information Processing and Retrieval, New York
[2]. A. Agarwala, P. Singh, and P. K. Atrey, "DICE: A dual integrity convergent encryption protocol for client side secure data deduplication," in *IEEE International Conference on Systems, Man, and Cybernetics*, Banff, Canada, 2017, pp.2176–2181.
[3]. M. Bellare, S. Keelveedhi, and T. Ristenpart, "Messagelocked encryption and secure deduplication," in *Advances in Cryptology – 32nd Annual International Conference on the Theory and Applications of Cryptographic Techniques*, Athens, Greece, 2013, pp. 296–312.
[4]. F. Rashid, A. Miri, and I. Woungang, "Secure image deduplication through image compression," *J. Inf. Secur. Appl.*vol. 27, no. C, pp. 54–64, 2016.
[5]. D. Li, C. Yang, C. Li, Q. Jiang, X. Chen, J. Ma, and J. Ren, "A client-based secure deduplication of multimedia data," in *IEEE International Conference on communications*, Paris, France, 2017, pp. 1–6
[6]. X. Li, J. Li, and F. Huang, "A secure cloud storage system supporting privacy-preserving fuzzy deduplication," *Soft Computing*, vol. 20, no. 4, pp. 1437–1448, 2016.
[7]. Mihir Bellare, Sriram Keelveedhi, Thomas Ristenpart, " Message-Locked Encryption and Secure Deduplication" Eurocrypt 2013.
[8]. M. Bellare and S. Keelveedhi, "Interactive message-locked encryption and secure deduplication," in *Public-Key Cryptography – 18th IACR International Conference on Practice and Theory in Public-Key Cryptography*, Gaithersburg, MD, USA, 2015, pp. 516–538.
[9]. E. Torres, G. Callou, G. Alves, J. Accioly, and H. Gustavo, "Storage services in private clouds: Analysis, performance and availability modeling," in *IEEE International Conference on Systems, Man, and Cybernetics*, Budapest, Hungary, 2016, pp. 3288–3293.

