International Journal of Scientific Engineering and Technology Research

ISSN 2319-8885
Vol.05,Issue.20
August-2016,
Pages:4143-4146

www.ijsetr.com

A Survey on Secure and Authorized De-Duplication using Hybrid Clouds

K. SWATHI[1], DR. N. KASIVISWANATH[2]

[1]PG Scholar, Dept of CSE, G.Pulla Reddy Engineering College (Autonomous), Kurnool, AP, India,
E-mail: swathi.muthya@gmail.com.

[2]Professor & HOD, Dept of CSE, G.Pulla Reddy Engineering College (Autonomous), Kurnool, AP, India.

Abstract: Nowadays, cloud computing provides very large amount of storage space and massive parallel computing at effective cost. It provides all kinds of services for the users. The main service offered by the cloud is storage facility. As cloud computing becomes popular because excessive amount of data can be stored in the cloud. One negative challenge of cloud storage space services is the managing of the ever-growing volume of data. However, exponential growth of ever-increasing volume of data has elevated many new challenges. Many Cloud storage service providers such as Dropbox, Mozy etc performs de-duplication to save space by only storing one copy of each file uploaded instead of storing Duplicates. Data deduplication is one of specialized data compression technique for eliminating duplicate copies of repeating data and has been extensively used in cloud storage to reduce the amount of storage space and save bandwidth. To protect the privacy of sensitive data while supporting deduplication, the content hash keying is an encryption method has been proposed to encrypt the data before outsourcing. To better protect data security, we present special privileges of user to address trouble of authorized data de-duplication.

Keywords: Deduplication, Authorized Duplicate Check, Confidentiality, Hybrid Cloud, Content Hash Keying.

I. INTRODUCTION

Cloud computing provides unlimited "virtualized" resources to users as services over the Internet, At the same time it hides the platform and implementation details. Today's cloud service providers offer both highly available storage and extremely parallel computing resources at fairly cheaper costs. As cloud computing becomes popular, an increasing amount of data is being stored in the cloud and shared by users with specific privileges, which defines the access rights of the stored data. One critical challenge is the management of the ever-increasing volume of data. To make data management scalable and to reduce the increased amount of data in cloud, de-duplication has been a well-known technique. Data de-duplication is a intelligent data compression technique for eliminating duplicate copies of repeating data in storage. This procedure is used to improve storage utilization and efficiency of cloud storage. De-duplication allows saving storage space and minimizing redundant data. In this model duplicate data is stored only once and we provide pointers to the actual data. Traditional encryption does not work because different users use their own keys to encrypt their data. Thus identical data copies of different users will lead to different cipher texts, making deduplication impossible. Deduplication can take place at two levels either the file level or the block level. For file-level deduplication, it discards the duplicate copies of the same file. Deduplication can also take place at the block level, which ignores the identical blocks of data that occur in non-identical files.

Although data deduplication brings a lot of benefits, but security and privacy concerns originate as user's sensitive data are susceptible to both inside and outside attacks. Content hash keying is an encryption technique used to combine the storage saving of de-duplication to enforce confidentiality. In Content hash keying encryption, the data copy is encrypted with a key derived by hashing the data itself then we get a convergent key. This convergent key is used for both encryption and decryption of a data copy. After key generation and data encryption, then users preserve the keys and send the cipher text to the cloud. Since encryption is deterministic, identical data copies will generate the similar convergent key and the same cipher text. This allows the cloud to perform de-duplication on the cipher texts. The cipher texts can only be decrypted by the corresponding data owners with their convergent keys. However, previous de-duplication systems cannot support differential authorization duplicate check, which is important in many applications. In such an authorized deduplication system, each user is issued a set of privileges during system initialization .Each file uploaded to the cloud is also bounded by a set of privileges to specify who are the right persons to perform the duplicate check and access the files. Before submitting his duplicate check request for some file, the user needs to take this file and his own privileges as inputs. The user is able to find a duplicate for this file if and only if there is a copy of this file and a matched privilege is already stored in the cloud.

II. RELATED WORKS

Cloud computing is an emerging technology in market. Day by day application hosting on cloud increases rapidly causes huge data storage on cloud most of them are duplicated. Due to this the main challenge faced by cloud service provider is the management of this ever-increasing bulk data.

S. Quinlan and S. Dorward. Venti [1] – In 2002, this paper presents an approach towards de-duplication called write-once policy of data. It provides efficient storage applications such as backup system i.e. logical backup, physical backup, and snapshot file systems.

J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer[2] – This paper introduces convergent key technique. To enforce data confidentiality and making de-duplication feasible convergent encryption is proposed. By applying cryptographic hash function on data convergent key is generated. Using this key It encrypts/decrypts a data. Encrypted data is sent to the cloud and user preserves the key and sends the cipher text to the cloud. The encryption is deterministic operation. The key is derived from the data content, hence identical data copies will generate the same convergent key and using the same key same cipher text is generated.

Pinkas, and A.Shulman-Peleg[3] – To prevent unauthorized access, a secure proof of ownership protocol is also needed to provide the proof that the user indeed owns the same file when a duplicate is found. After the proof, subsequent users with the same file will be provided a pointer from the server without needing to upload the same file. A user can download the encrypted file with the pointer from the server, which can only be decrypted by the corresponding data owners with their convergent keys.

M. Bellare, S. Keelveedhi, and T. Ristenpart[4] –Message-locked encryption and secure de-duplication: In this they formalize a new cryptographic primitive that they call Message-Locked Encryption (MLE), where the key under which encryption and decryption are performed is itself derived from the message. MLE provides a way to achieve secure de-duplication, a goal currently targeted by numerous cloud storage providers.

Weak leakage-resilient client-side de-duplication of encrypted data in cloud storage by Xu et al.[5]- also addressed the problem and showed a secure convergent encryption for efficient encryption.The proposed technique only focuses on encryption and file level de-duplication. The issue of key-management and block-level de-duplication is not considered.

D. Ferraiolo and R. Kuhn. [6] – Role-based access controls: In this they represent limitation of Mandatory Access Controls (MAC) technique. This is required for high level security like multilevel secure military applications.

Architecture for secure cloud computing - Bugiel et al. [8] – It provided an architecture consisting of twin clouds for securely outsourcing of user private data and arbitrary computations to an untrusted commodity cloud.

Zhang et [9] al also presented the hybrid cloud techniques to support privacy-aware data-intensive computing. We consider addressing the authorized privileged de-duplication problem over data in public cloud. The security model of our systems is similar to those related work, where the private cloud is assume to be honest but curious.

S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-[10] – Proposes of POW (proof of ownership) technique is that a user can efficiently prove to the cloud storage server that he/she owns a file without uploading the file itself. It also proposes the Merkle-Hash Tree to enable client-side de-duplication, which include the bounded leakage setting. The proposed scheme is focusing only on the data ownership and not on the data privacy.

In S. Ossowski and P. Lecca: [11]- extended proofs of ownership mechanism for encrypted files. These papers do not address how to minimize the key management overhead.

Pietro and Sorniotti [14]- proposed another efficient PoW scheme by choosing the projection of a file onto some randomly selected bit-positions as the file proof But this project do not deal with data privacy.

III. SYSTEM MODEL

A. Contributions

Aiming at efficiently solving the problem of de-duplication with differential privileges in cloud storage, we consider a Twin cloud architecture consisting of a public cloud and a private cloud. Unlike existing data deduplication systems, the private cloud is involved as a proxy to permit data owner/users to securely perform duplicate check with differential privileges. This architecture is practical and has attracted much attention from researchers. The data owners only outsource their data by utilizing public cloud while the data operations are managed in private cloud. A new de-duplication system supporting differential privileges duplicate check is proposed under this Twin cloud architecture where the S-CSP resides in the public cloud. The user is only permitted to perform the duplicate check for files marked with the corresponding privileges.

B. Preliminaries

- Symmetric encryption
- Content hash keying encryption
- proprietorship protocol
- Identification protocol

Here, we address the problem of privacy preserving de-duplication in cloud computing and propose a new de-duplication system that includes the public cloud and the private cloud known as hybrid cloud which is a combination of the both public cloud and private cloud. In general if we

used the public cloud we can't provide the security to our private data and hence our private data will be loss. So that we have to provide the security to our data for that we make a use of private cloud also. In this system we also provide the data de-duplication, which is used to avoid the duplicate copies of data. User can upload and download the files from public cloud but private cloud provide the security for that data. That means only the authorized persons can upload and download the files from the public cloud, for that user generates the key and stored that key onto the private cloud. At the time of downloading user request to the private cloud for key and then access that Particular file.

C. Methodology

If the user wants to upload the file on the public cloud then user first has to encrypt that file with the convergent key and then send it to the public cloud, at the same time user also generates the key for that file and sends that key to the private cloud for the reason of security. In the public cloud we use one algorithm for de-duplication. Which is used to keep away from the duplicate copies of files which are entered in the public cloud. Hence it also minimizes the bandwidth, which means we require the less space for storing the files on the public cloud. In the public cloud any person who is unauthorized can also access or store the data so we can conclude that in the public cloud the security is not provided. In order to provide more security user can use the private cloud instead of using the public cloud. User generates the key at the time of uploading file and store it to the private cloud. When user wants to downloads the file that he/she upload, he/she sends the request to the public cloud. Public cloud provides the list of files that are uploaded by the many users because there is no security is provided. When user select one of the file from the list of files then private cloud sends a message like enter the key.User has to enter the key that has generated for that file. When user enter the key the private cloud checks the key for that file and if the key is correct then that user is valid, now private cloud give access to that user to download that file successfully, then user downloads the file from the public cloud and decrypt that file by using the convergent key which is used at the time of encryption of that file.

IV. SYSTEM ARCHITECTURE

In our architecture there are three modules

- User
- Public cloud
- Private cloud

S-CSP: The purpose of this entity to work as a data storage service in public cloud. The S-CSP eliminates the duplicate data using de-duplication technique and keeps the unique data as it is. S-SCP entity is used to reduce the storage cost. S-CSP has abundant storage capacity and computational power.

Data User: A user is an entity that wants to access the data or files from S-SCP. User generate the key and store that key in private cloud. In storage system supporting de-duplication, the user uploads unique data but do not upload any duplicate data to save the upload bandwidth, which may be owned by the same user or different users as shown in Fig.1. Each file is protected by convergent encryption key and can access by only authorized persons. In our system user must need to register in private cloud for storing token with respective file which are stored on public cloud. When he wants to access that file he access respective token from private cloud and then access his files from public cloud.

Private Cloud: In general for providing more security user can use the private cloud instead of public cloud. User store the generated key in private cloud.At the time of downloading system asks the key to download the file. User cannot store the secret key internally. For providing proper protection to key we use private cloud. Private cloud only stores the convergent key with respective file. When user wants to access the key it first checks authority of user then and then provides a key.

Public Cloud: Public cloud is an entity used for the storage purpose .User upload the files in public cloud. Public cloud is similar as S-CSP.When the user wants to download the files from public cloud, it will be asking the key which is generated or stored in private cloud. When the users key is match with files key at that time user can download the file, without key user cannot access the file. Only authorized user can access the file. In public cloud all files are stored in encrypted format. If any chance unauthorized person hack our file, but without the secret or convergent key he doesn't access original file. On public cloud there are lots of files are stored each user access its respective file if its token matches with S-CSP server token.

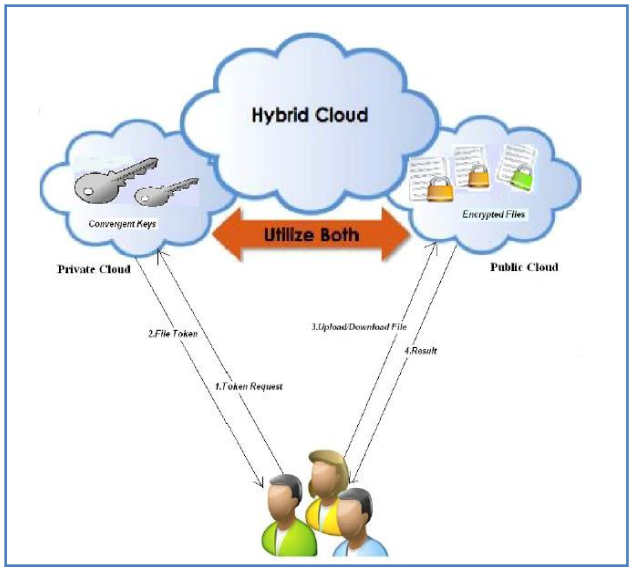

Fig.1. Architecture of Authorized Deduplication.

A. Operations Performed On Hybrid Cloud

File Uploading: When user want to upload the file to the public cloud then user first encrypt the file which is to be upload by making use of the symmetric key and send it to the Public cloud. At the same time user generates the key for that

file and sends it to the private cloud. In this way user can upload the file into the public cloud.

File Downloading: When user wants to download the file that he/she has uploaded on the public cloud. he/she make a request to the public cloud, then public cloud provide a list of files that many users uploaded on it. Among that user select one of the file from the list of files and enter the download option at that time private cloud sends a message that enter the key for the file generated by the user, then user enters the key for the file that he/she is generated, then private cloud checks the key for that file and if the key is correct that means the user is valid. Only then the user can download the file from the public cloud. Otherwise user can't download the file. When user download the file from the public cloud it is in the encrypted format then user decrypt that file by using the same symmetric key.

V. CONCLUSION

The idea of authorized data deduplication was proposed to protect the data security by including differential authority of users in the duplicate check. In public cloud our data are securely store in encrypted format, and also in private cloud our key is store with respective file. There is no need to user remember the key. So without key anyone can not access our file or data from public cloud. This paper will provide more efficiency and security in cloud computing using authorized deduplication check and hierarchical access control method. It will improve the storage efficiency and performance of cloud storage.

VI. REFERENCES

[1] M. Bellare, S. Keelveedhi, and T. Ristenpart. "Dupless: Server aided encryption for deduplicated storage." In USENIX Security Symposium, 2013.

[2] J. Li, X. Chen, M. Li, J. Li, P. Lee, and W. Lou. Secure deduplication with efficient and reliable convergent key management. In IEEE Transactions on Parallel and Distributed Systems, 2013.

[3]W. K. Ng, Y. Wen, and H. Zhu. "Private data de-duplication protocols in cloud storage." In S. Ossowski and P. Lecca, editors, Proceedings of the 27th Annual ACM Symposium on Applied Computing, pages 441–446. ACM, 2012.

[4] R. D. Pietro and A. Sorniotti. "Boosting efficiency and security in proof of ownership for de-duplication." In H. Y. Youm and Y. Won, editors, ACM Symposium on Information, Computer and Communications Security, pages 81–82. ACM, 2012.

[5] S. Bugiel, S. Nurnberger, A. Sadeghi, and T. Schneider. "Twin clouds: An architecture for secure cloud computing." In Workshop on Cryptography and Security in Clouds (WCSC 2011), 2011.

[6] K. Zhang, X. Zhou, Y. Chen, X.Wang, and Y. Ruan. "Sedic: privacyaware data intensive computing on hybrid clouds.", In Proceedings of the 18th ACM conference on Computer and communications security, CCS'11, pages 515–526, New York, NY, USA, 2011. ACM.

[7] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg. "Proofs of ownership in remote storage systems." In Y. Chen, G. Danezis, and V. Shmatikov, editors, ACM Conference on Computer and Communications Security, pages 491–500. ACM, 2011.

[8] Pinkas, and A. Shulman-Peleg. "Proofs of ownership in remote storage systems." In Y. Chen, G. Danezis, and V. Shmatikov, editors, ACM Conference on Computer and Communications Security, pages 491–500. ACM, 2011

[9] S. Quinlan and S. Dorward. Venti:"a new approach to archival storage", In Proc. USENIX FAST, Jan 2002

[10] J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer. "Eclaiming space from duplicate files in a server less distributed file system.", In ICDCS, pages 617–624, 2002. S. Halevi, D. Harnik, B.

[11] K. Zhang, X. Zhou, Y. Chen, X.Wang, and Y. Ruan. "Sedic: privacyaware data intensive computing on hybrid clouds.", In Proceedings of the 18th ACM conference on Computer and communications security, CCS'11, pages 515–526, New York, NY, USA, 2011. ACM.