International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 03 Issue 17
November 2016

Online-Storage Data Auditing and Secure Source Side Deduplication

PONNADA BHAVYA BHASKARI[1] ,REDDYBOINA ASHOK[2]
[1]PG Scholar, Dept of CSE,Vasireddy Venkatadri Institute of Technology,Nambur,Guntur,AP,India,
E-mail:bhavya.bhaskari@gmail.com
[2]Assistant Professor,Dept of CSE,Vasireddy Venkatadri Institute if Technology,Nambur,Guntur,AP,India,
E-mail:anshok04@gmail.com

Abstract:
Online-Storage computing is an common data interactive paradigm for processing of large amounts of data and storage without consider the local infrastructure limitations. The advent of Online-Storage storage enables the organizations and enterprises to outsource their data to third party Online-Storage service providers (CSP). Though the services provided by Online-Storage has many advantages, the users willingly give up the physical control of their outsourced information which inevitably poses new security and privacy risks and yet another confront is the organization of ever rising volume of data for CSP. In order to compact with these safety issues, a new safe storage with deduplication scheme has been adopted. In order to give security of outsourced data beside hateful users with snooping CSP's, a new convergent encryption method is proposed. Every Client encrypts the file previous to uploading and the encrypted file is enter to Hash algorithm which generate a single identifier for all file. The Client specify the official users and their admission rights in a metafile uploaded to the Online-Storage and user can decrypt the downloaded file with his secret key in.

Keywords: Online-Storage Storage, Deduplication, Integrity, Security, Privacy.

1.INTRODUCTION

Online-Storage Computing is defined as "A large-scale spread computing paradigm that is driven by economies of scale, in which a group of abstracted, virtualized, dynamically scalable, managed computing power, storage space, platform, and services are delivered on demand to outside customers over the Internet[1]".It is the means of delivering any and all information technology components from computing power to computing infrastructure, application ,business processes and collaboration actually deliver IT as a service. Online-Storage computing be an advanced way of computing where applications, data and resources are provide to user as a service over the web .

Moving the data onto the Online-Storage offers significant advantages in resource saving and provides great ease to users as they don't have to worry about the difficulty of hardware, software and their maintenance. With the potentially unlimited storage space offered by Online-Storage providers, users tend to use as a lot as space as they can and vendors continuously look for techniques aimed to minimize redundant data and maximize gap savings. .Technology is changing every day and organizations are expected to adopt to the changes and transform enterprise IT with self-service, charge back, service catalogs, resource orchestration, total application provisioning hybrid IT, reservations, etc. A technique which has been commonly used and adopted is deduplication. Deduplication is a method that stores only a only copy of each file on a storage server not considering of how several clients invite to store that file.

Despite the important advantages, it brings several new security issue towards the user's outsourced data and privacy is assured by encryption. unluckily, deduplication and encryption are two contradictory technologies. While the aim of deduplication is to detect the same data segments plus preserve them only one time, the effect of encryption is to make two the same data segments indistinguishable after being encrypted. This way that if data is encrypted in a standard way Online-Storage, the Online-Storage storage provider cannot apply deduplication since two same data segments will be different after encryption. On the additional hand if data is not encrypted by user, confidentiality cannot be definite and data are not protected against curious Online-Storage storage providers.

In order to gather the two conflicting requirements a technique that is convergent encryption have been planned in which the data has to be uploaded is encrypted with the key generate from the hash of its contents. Convergent encryption is a high-quality candidate to get confidentiality and deduplication at the similar time. The safety of the this system relies on its new architecture where in adding to basic storage provider; a metadata manager is defined. The privacy is achieved through convergent encryption and the metadata manager is in charge for key management task. Thus, the original deduplication is done at the file level and at client side.

The remainder of the work is planned as follows. First in section II a brief background on deduplication and convergent encryption have been explained. Section III provide an summary of the connected work. Section IV describe the Online-Storage storage architectures used. Section V describes the future

system and then followed by security analysis. lastly Section VII present conclusion and future work.

II.BACKGROUND

A. Deduplication

reference that points toward the store chunk.

There are mostly two categories of: file-level deduplication and block-level deduplication. In block level deduplication the block size can also be permanent or variable. It is also categorized base on the location at which the deduplication is done: if the data is deduplicated at the server side then it is call target based deduplication or else source-based deduplication. In target-based deduplication the file/data is primary sent to the server and then the deduplication is done while in source-based deduplication, the client hashes the data to be uploaded and sends the unique hash value to the Online-Storage server to check for individuality of the data. While deduplication on the client side can achieve bandwidth savings, it unluckily can make the system open to to side-channel attacks. On the extra hand, by deduplicating data at the storage server, the system is protected against side channel attacks but the does not decreases the communication overhead.

B. Convergent Encryption

The fundamental idea of convergent encryption (CE) is to get the encryption key from the hash of the plaintext. The easiest implementation of convergent encryption can be done as follows: Alice derive the encryption key from her message such as K = H (M), where H is cryptographic hash function; she can encrypt the message with this key, therefore: C = E (K; M) = E (H (M); M), where E is block cipher. By using this technique, two users by two same plaintexts will obtain two same cipher texts since the encryption key is the same; so the Online-Storage storage provider can be able to perform deduplication on that cipher texts. Furthermore, encryption keys are generated, retain and protected by users. since the encryption key is deterministically generated from the plaintext, users do not has to relate with each other for establish an agreement on the key to encrypt a given plaintext. Therefore, convergent encryption seem to be a good quality candidate for the adoption of encryption and deduplication in Online-Storage storage domain.

III.RELATED WORK

Many systems has been developed to give secure storage space but traditional encryption processes are not proper for deduplication process. Most works do not consider security as a concern for deduplicating systems. However Zhifengxiao et al. [1] systematically considered the security and privacy challenge in Online-Storage computing environment based on attribute driven methodology. The authors recognized the most representative safety/privacy attributes and the vulnerabilities, which may be exploited by adversaries in order to do various attacks and a few of the defense strategy were also discussed.

Wenjing Lou et al. [2] focused on the Online-Storage data storage security. The authors proposed an efficient and flexible scheme by using the homomorphism token with distributed confirmation of erasure-code data which gets the integration of storage correctness cover and data error localization.

Cond Wang et al. [3] developed a flexible distributed storage space integrity audit mechanism which allows the user to audit the Online-Storage storage with lightweight communication and computation cost. The auditing result not only ensure strong Online-Storage storage correctness guarante, but also simultaneously achieves fast error localization.

Luca Ferretti et al. [4] designed a novel architecture so as to integrates Online-Storage database services by means of data confidentiality and possibility of performing concurrent operations on encrypted Online-Storage database and it also eliminate the intermediate proxies that bound the elasticity, availability and scalability properties that be intrinsic in Online-Storage-based solutions.

Hong Liu et al. [5] implemented a shared power base privacy-preserving authentication procedure to address the privacy issues of Online-Storage storage; the proposed protocol is attractive for multi-user joint Online-Storage applications. KanYag et al. [6] proposed the Cipher text-Policy Attribute-based Encryption (CP-ABE) to manage the data access in Online-Storage storage. The authors planned the efficient and revocable data access manage scheme for multi-authority Online-Storage storage systems. The method described achieved both forward and backward security

Douceur et al. [7] study the trouble of Deduplication in multi-tenant environment. The authors planned the use of convergent encryption i.e., deriving keys from the hash of plain text with attempt to join data confidentiality with the possibility of data deduplication. Then storer et al. [8] pointed out some security problem and obtained a security model for secure data Deduplication. though, the two protocols concentrate on server-side Deduplication and do not reflect on data leakage settings, against malicious users.

Halevi et al. [9] the concept of proof of ownership (POW) was introduce in order to avoid the private data leakage. These scheme involves the server challenging the client to present the suitable sibling paths for a subset of a merkle tree leaves.

Ng et al. [10] proposed a POW method on encrypted data. The file is divided into fixed-size blocks, where each block has a single commitment.

In Deduplication procedure, unique chunks of data or byte pattern are recognized and store during analysis. As the analysis continue, other chunks are compare to the stored copy and whenever a match occurs, the unneeded chunk is replace with a little

International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 03 Issue 17
November 2016

Hence, the owner have to prove the possession of data chunk of precise commitment, with no need to reveal the secret information. on the otherhand this scheme introduces a high computation cost.

IV.ONLINE-STORAGE ARCHITECTURE

A. Architecture

Fig. 1 illustrates the descriptive network architecture of Online-Storage storage. It relies the following entities for good management for client data.

- **Online-Storage Service Provider (CSP)***:* a CSP have significant resources to rule distributed Online-Storage storage server and to handle its database servers. It also provides virtual infrastructure to host application services. These services can be used by client to manage his datastored in the Online-Storage servers.

- **Client:** a client makes use of provider's resources to store, retrieve and share data with many users. A client can either an individual or an enterprise

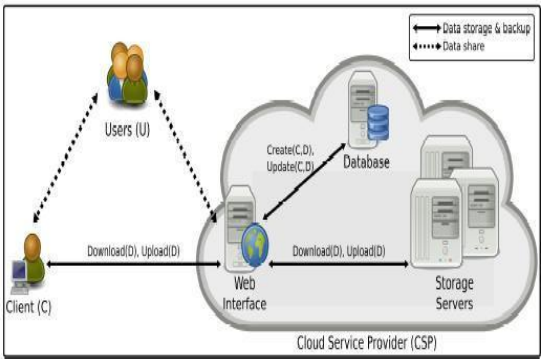

Fig 1 Architecture of Online-Storage data storage

- **Users:** the users are able to use the content stored in the Online-Storage, based on their access rights which are authorizations granted by the client, like the rights to read, write or re-store the modified data in Online-Storage.

In practice, the CSP provides a web interface for the client to store data into a set of Online-Storage servers, which are running a cooperated and distributed manner. In adding, the web interface is used by the users to retrieve, modify and restore data from the Online-Storage, depending on access rights. Moreover, the CSP relies on database servers to combine client identities to their store data identifiers and group identifiers.

B. Security Requirements

When outsourcing data to a third party, provided that confidentiality and privacy become more challenging and conflicting.

Privacy is a serious concern with regard to Online-Storage storage due to the fact that client data reside between distributed public servers. thus, there are potential risks where the private information (e.g., financial data, health record) or personal information (e.g., personal profile) is disclosed. Meanwhile, confidentiality imply that client's data have should be kept secret from both Online-Storage provider and other users.

Confidentiality remains a few of the greatest concerns.Thisislargelyduetothefactthatusersoutsourceth eirdataonOnline-Storageservers, which are controlled and managed by potentially untrusted CSPs. That is why, it is compulsory to provide secrecy by encrypting data previous to their storage in Online-Storage servers when keeping the decryption keys out of the reach of CSP and any malicious user. For designing the most suitable security solutions for Online-Storage storage, we are considering an honest but curious Online-Storage provider, as a threat model. That is, it honestly performs the operations defined by our planned scheme, but it may actively attempt to gain the knowledge of outsourced data. In adding, an attacker can be either be

C. Assumptions

Our solution consider the following assumption .First; we assume that there is an established secure channel among the client and the CSP. This secure channel supports mutual authentication and data privacy and integrity. Hence, after effectively authenticating with the CSP, these Online-Storage user share the same resources in multi-tenant environment.

Second, our solution use the hash function in the generation of the enciphering data keys. Hence, we assumes that these cryptographic functions are powerfully collision resistant, as it is an intractable problem to find the same output for unlike data files.

V.PROPOSAL FOR SECURE DATA STORAGE, BACKUP, SHARING

The proposed scheme performs Client side deduplication and it is based on three different scenarios: Storage Backup, and sharing schemes.

A. Online-Storage Data Storage

When a client wants to store a new data file f on the Online-Storage, he derives the enciphering key kf from the data contents, based on one-way hash function H (). Note that data be stored enciphered during Online-Storage servers, based on symmetric algorithm. Hence, the data owner has to encipher the data, file that he intends to outsource. Then, it generates the data identifier ID. That is, it is the Hash on encrypted data. This identifier, associated to the file, must be unique in the CSP database.

Hence, the client starts the storage process by

sending ClientRequestVerif message to verify the uniqueness of the generated ID to his CSP.

New Data File Storage: The storage procedure consists in exchanging the four following messages:

- **ClientRequestVerif:** this first message contains the generates a data identifier ID. This message is a request for the proof of the uniqueness of the ID. The CSP replies with a ResponseVerif message to validate or invalidate the claimed identifier.

- **ResponseVerif:** the acknowledgement message is generated by the CSP to notify the client about the existence of the requested MTf in the database.

and backward secrecy. **ClientRequestStorage:** This message is send by the client. If the file does not be present in the Online-Storage servers, the client sends the file that he intends to store in the Online-Storage, with data decrypting key kf enciphered with the public keys of approved users. Then, the enciphered kf is included in Meta data of the file and it serves as an access rights provision.

- **ResponseStorage:** This acknowledgement message, sent by CSP, is used to verify to the client the success of his data storage.

B. Online-Storage Data Backup

The data backup process start when the client request for retrieving the data previously present in the Online-Storage. The data backup process include the following messages:

- **ClientRequestBackup:** It contains the URL of the requested data that the client wants to retrieve. After receiving this client requests, the CSP verifies the client ownership of the claimed file and generate a ResponseBackup message.
- **ResponseBackup:** In his reply, the CSP include the encrypted outsourced data kf (f).After receiving the ResponseBackup message, the client initially retrieve the file metadata and deciphers the data decrypting key kf , with the secret key. Then, he uses derived key to decrypt the request data file.

C. Online-Storage Data Sharing

In data sharing process, the client outsources his data to the Online-Storage and authorizes a set of users to access the data. Users' access rights are given by data owner and managed by CSP. That is, these access rights are also includes the metadata file. In totaling, the CSP is in charge of verify each recipient access permissions before sending him outsourced data.

Each member in the group can start the data sharing process base on the two following messages:

- **UserRequestAccess***:* This message contains the URL to the requested file. When receiving this message, the CSP search for the read/write permissions of receiver, and then, he generates a Response Access message.

- **ResponseAccess:** The CSP includes, in its response, the enciphered file kf (f). Upon receiving this message, all recipients retrieves the data decrypting key from user metadata. so as to, he deciphers the associated symmetric key with his own private key. Then,he performs a symmetric decryption algorithm to get back the plaintext.

Our proposal provides a strong solution to improve the confidentiality of data to the Online-Storage. In addition, the access to outsourced data is restricted by two processes. First, there is traditional access list managed by CSP. Secondly, the client has to hold the private decrypting key to get the secret needed to retrieve the symmetric key compulsory needed to decipher data.

VI.SECURITY DISCUSSION

In this section the informal security analysis to the proposal is discussed. In addition, the possible refinements that could be made to mitigate other threats are also discussed.

Data confidentiality – In the present model, it is proposed to outsource encrypted data to remote storage servers. That,the data is stored enciphered in the Online-Storage, based on a symmetric encryption algorithm using a per data key. This enciphering key is content linked information, ensuring data deduplication in remote servers. Thus, the confidentiality of outsourced data is twofold. First; we make sure confidentiality preservation against malicious users. On one hand, when a user wants to store new data in Online-Storage, he has to send the data identifier ID, based on encrypted file. Hence, this dual data identifier protection provides improved secrecy to data outsourcing issue.

Second, we enhance data confidentiality against curious servers. That is, the owner of data outsources encrypted contents. Then, he enciphers the decrypting key relying on asymmetric scheme, in order to ensure efficient access control. As such, the CSP is also unable to learn the contents of stored data in the public servers.

Privacy – Based on a cryptographic solution to put data content secret, sensitive information are usually built-in in metadata whose leakage is critical concern in a multi-tenant environment. Thus, our model mitigates to such privacy violation issue. On one side, the CSP identifies clients as data owners, when outsourcing the similar content of remote servers. However, the Online-

Storage server cannot combine the consistency between the plaintext information and these data owners, as he have only access to hashed identifiers and encrypted contents. Consequently, he is unable to build user profiles, based on received identifiers.

VII.CONCLUSIONS

The growing want for secure Online-Storage storage services and attractive properties of cryptography lead to the innovative solution to the data outsourcing security issue.

Our solution is based on a cryptographic usage of symmetric encryption used for enciphering the data file and asymmetric encryption for Meta data files, due to highest sensibility of this information towards several intrusions. Besides, our solution is also shown to be resistant to unauthorized access to data and to any data disclosure during sharing process, providing two levels of access control verification. Finally, we believe the Online-Storage data storage security is still full of challenges and of chief importance, and many research problems remain to be identified.

REFERENCES

[1] Zhifeng Xiao, Yang Xiao, *Security and Privacy in Computing*, IEEE Communications Surveys and &Tutorials, Vol. 15, No. 2, pp. 843- 859, Second quarter 2013.

[2] Cong Wang, Qian Wang, KuiRen, Wenjing Lou, *Ensuring Data Storage in Online-Storage Computing*, in Proc. Of IWQoS'09, pp. 1-9, July 2009.

[3] Cong Wang, Qian Wang, KuiRen, Wenjing Lou, *Towards Secure and Dependable Storage Services in Online-Storage Computing*, IEEE Transactions in Services Computing , Vol. 5, No. 2, pp.220-232, 2012.

[4] Luca Ferreti, Michele Colajanni, MircoMarchetti, *Distributed, Concurrent and Independent Access to Encrypted Online-Storage Databases*, IEEE Transactions in Parallel and Distributed Systems, Vol. 25, No. 2, pp. 437-446, Feb 2014.

[5] Hing Liu, HuanshengNing, QingxuXiong, Laurence T. Yang, *Shared Authority Bassed Privacy-preserving Authentication Protocol in Online-Storage Computing*, IEEE Transactions in Parallel and Distributed Systems, Vol: pp:99, 2014.

[6] J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon an M. Theimer, *Reclaiming Space from Duplicate files in a Serviceless Distributed File System*, in Proc. of 22nd International Conference on Distributed Computing Systems (ICDCS-2012).

[7] M. Dutch, *Understanding Data Duplication Ratios*, SNIA White Paper, June 2008.

[8] T. G. et.al, *GNU Multiple Precision Arithmetic Library*, 4.1.2, December 2002.

[9] S. Halevi, D. Harnik, B. Pinas, A. Shulman-Peleg, *Proofs of Ownership in Remote Storage Systems*, in Proc. Of the 18th ACM Conference on Computer and Communications Security, CCS'11, pp-491-500, New York, NY, USA, 2011.

[10] Danny Harnik, Benny Pinkas and Alexander Shulman-Peleg, Side,*Channels in Online-Storage Services: Deduplication in Online-Storage Storage, Security and Privacy*, IEEE, 8(6):40-47, 2010.

[11] Mihir Bellare, AlexandraBoldyreva and Adam ONeill.Deterministic and efficiently searchable encryption.In Advances in Cryptology-CRYPTO 2007,pages 535-552.Springer,2007.

Author's Profile:
PONNADA BHAVYA BHASKARI is pursing M.Tech (CSE) from Vasireddy Venkatadri Institute of Technology,Nambur, Guntur.

REDDYBONIA ASHOK is working as assistant professor in CSE department of Vasireddy Venkatadri Institute of Technology, Nambur, Guntur.He received M,Tech(CSE) from Pondicherry university ,Pondicherry . His Research interests are in the areas of Data Mining, Big Data Analytics and Natural Language Processing.