ISSN 2395-1621

De-Duplication & Secure Authorized Data Using Hybrid Cloud

[#1]Nimgire Reshma, [#2]Kemdarne Snehal, [#3]Tambe Tejas, [#4]Prof. G. M. Bhandari

[1]reshmanimgire7@gmail.com
[2]sneha.kemdarne@gmail.com
[3]tejastambe9@gmail.com

[#123]Student, Department of Computer
[#4]HOD, Department of Computer
JSPM's
BSIOTR, Wagholi,Pune.

ABSTRACT

In recent days, Cloud storage systems are becoming increasingly popular which provides highly available storage. Cloud storage service providers such as Dropbox, Amazon and others perform deduplication to save space by only storing one copy of each file uploaded. Data deduplication is one of the important data compression techniques. Data deduplication technique is used to eliminate duplicate data copies by keeping a physical copy. To reduce storage space and save bandwidth in cloud storage data deduplication is used. For that purpose convergent encryption has been extensively adopt for secure deduplication. To retain the confidentiality of sensitive data while supporting the deduplication, convergent cryptography technique is used to cipher the data before contract out. In hybrid cloud architecture, several new deduplication construction supporting authorized duplicate check also present. To improve scalability and storage problem data deduplication is used. The proposed security models contain the illustration of security analysis scheme. In proposed system authorized duplicate check scheme is used to incur minimal overhead compared to normal operations.

Keywords: Deduplication, confidentiality, Authorized duplicate check, Convergent encryption technique.

ARTICLE INFO

Article History

Received:18th October 2015

Received in revised form :20th October 2015

Accepted :22nd October 2015

Published online :24th Octber 2015

I. INTRODUCTION

Cloud computing provides apparently unrestricted "virtu-alized" resources to users as services through the whole Internet. Also hide platform & implementation de¬tails from cloud. Cloud service providers provide both highly available storage and comparatively low costs for parallel computing. Now cloud computing becomes common, increasing amount of data is being stored in the cloud which is shared by users with some different privileges, which define the access rights to particular user. To manage the increasing volume of data is critical challenge of cloud storage services. To make data formation scalable in cloud computation, deduplication [17] is a well-known technique and has attracted more and more attending Recent. Data deduplication is a specialized data compression technique for eliminating duplicate copies of repetition data in storage. The technique is used to modify storage utilization and can also be applied to network data movement to trim the amount of bytes that must be sent. Instead of keeping multiple data copies with the same content, deduplicationobviate extra data by safekeeping only one physical copy. Deduplication can take place at either the file level or block level. For data file level deduplication, it eliminates duplicate copies of the same file. & for block level, which get rid of duplicate blocks of data that occur in non-identical file. Data deduplication is also provided the security and privacy. It concerns with users sensitive data are susceptible to both insider and outsider attacks. In traditional encryption technique, data confidentiality is provided but it is incompatible with data deduplication. Also, traditional encryption requires different users to encrypt their data with their own keys. Thus, identical data copies of various users will lead to various cipher text, which makes deduplication impossible. Convergent encryption is proposed to enforce data confidentiality while making deduplication feasible. It encrypts or decrypts a data copy with a merging key, which is incur by computing the cryptographic hash value of the content of the data copy. However, previous deduplication 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 which kind of users is allowed 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 stored in cloud. For example, in a company, many different privileges will be assigned to employees. In order to save cost and efficiently management, the data will be moved to the storage server provider (SCSP) in the public cloud with specified privileges and the deduplication technique will be applied to store only one copy of the same file. Because of privacy consideration, some files will be encrypted and allowed the duplicate check by employees with specified privileges to realize the access control. Traditional deduplication systems based on convergent encryption, although providing confidentiality to some extent; do not support the duplicate check with differential privileges. In other words, no differential privileges have been considered in the deduplication based on convergent encryption technique. It seems to be contradicted if we want to realize both deduplication and differential authorization duplicate check at the same time.

II. MOTIVATION

Cloud storage services are becoming very popular now a day. Cloud provides a better way of storage with efficient cost. One major problem with cloud is to manage huge amount of data. In order to manage data de-duplication technique is used. Although, de-duplication has many advantages but it has some security issues. This motivates us to propose a model which manage the security issues of de-duplication and provide authorized de-duplication in cloud.

III.NOTATIONS & PRELIMINARIES

Acronym	Description
S-CSP	Storage-cloud service provider
PoW	Proof Of Ownership
(pk_U,sk_U)	User's public and secret key pair
k_F	Convergent encryption key for file F
P_U	Privilege set of a user U
P_F	Specified privilege set of a fie F
$\Phi'_{F,p}$	Token of file F with privilege p

TABLE I

NOTATIONS USED

Symmetric encryption:
Symmetric encryption uses a common secret key κ to encrypt and decrypt information. A symmetric encryption scheme consists of three primitive functions:

- $KeyGen_{SE}(1^{\lambda})$ - k is the key generation algorithm that generates k using security parameter 1^{λ}
- $Enc_{SE}(k,M)$ - C is the symmetric encryption algorithm that takes the secret k and message M and then outputs the cipher text C
- DecSE(k,C) - M is the symmetric decryption algorithm that takes the secret k and cipher text C and then outputs the original message M.

Convergent encryption:
Convergent encryption [4], [8] provides data confidentiality in deduplication. A user or data owner derives a convergent key from each original data copy and encrypts the data copy with the convergent key. In addition, the user also derives a tag for the data copy, such that the tag will be used to detect duplicates. Here, we assume that the tag correctness property [4], if two data copies are same, then their tags are same. To detect duplicates, the user first sends the tag to the server side to check if the identical copy has been already stored. Both the encrypted data copy and its corresponding tag will be stored on the server side. Convergent encryption scheme can be defined with four primitive functions:

- $KeyGen_{CE}(M)$ - K is the key generation algorithm that maps a data copy M to a convergent key K
- $Enc_{CE}(K,M)$ - C is the symmetric encryption algorithm that takes both the convergent key K and the data copy M as inputs & then outputs a cipher text C
- $Dec_{CE}(K,C)$ - M is the decryption algorithm that takes both the cipher text C & the convergent key K as inputs and then outputs the original data copy M
- TagGen(M) - T(M) is the tag generation algorithm that maps the original data copy M and outputs a tag T(M).

Proof of ownership:
The use of proof of ownership(PoW) [11] enables users to prove their ownership ofdata copies to the storage server. Specifically, PoW isimplemented as an interactive algorithm (denoted byPoW) run by a user and a storage server.The formal security definitionfor PoW roughly follows the threat model in a contentdistribution network, where an attacker does not knowthe entire file, but has accomplices who have the file. The accomplices follow the "bounded retrieval model", suchthat they can help the attacker obtain the file, subject tothe constraint that they must send fewer bits than theinitial min-entropy of the file to the attacker [11].

Identification Protocol:
An identification protocol π can be described with two phases: Proof and Verify. In the stage of Proof, a prover/user U can demonstrate his identity to a verifier by performing some identification proof related to his identity. The input of the prover/user is his private keythat is sensitive information such as private key of a public key in his credit card number that he would not like to share with the other users. The verifier performs the verification with input of public information. At the conclusion of the protocol, the verifier outputs either accept or reject to denote whether the proof is passed or not.

IV. EXISTING SYSTEM

In existing system, the private cloud is involved as a proxy to allow data owner to securely perform duplicate check with different privileges.It requires extra bandwidth & more storage space. The data owners only outsource their data storage by utilizing public cloud while the data operation is managed in private cloud.
Drawbacks of existing system are User's sensitive data are susceptible to both insider and outsider attacks. Security is not provided in existing system. Management of the increasing volume of data. It requires extra bandwidth.

V. PROPOSED SYSTEM

We propose another advanced deduplication system supportingauthorized duplicate check. In this new deduplicationsystem, hybrid cloud architecture is used. The private keys for privileges willnot be issueddirectly to users, which will be kept andmanaged by the private cloud server. In this way,the users cannot share these private keys of privilegesin this proposed construction, which means that it canprevent the privilege key sharing among users in theabove straightforward construction. To get a file token,the user needs to send a request to the private cloudserver. The intuition of this construction can be describedas follows. To perform the duplicate check for some file,the user needs to get the file token from the privatecloud server. The private cloud server will also checkthe user's identity before issuing the corresponding filetoken to the user. Theauthorized duplicate check forthis file can be performed by the user with the publiccloud before uploading this file. Based on the results ofduplicate check, the user either uploads this file or runsPoW.

System Model:

Now we see the architecture of our system,

- S-CSP. This is an entity that provides a data storage service in public cloud. The S-CSP provides the data outsourcing service and stores data on behalf of the users. To reduce the storage cost, the S-CSP eliminates the storage of redundant data via deduplicationand keeps only unique data.
- User : A cloud user is which who wants to outsource data on public storage which acts as a public cloud in cloud computing. A system provides authenticate used to enter in system upload data with particular set of privileges for further accessing the uploaded data to download.
- Private cloud :Private cloud is able to provide data user/owner with an execution environment and infrastructure working as an interface between user and the public cloud. The private keys for the privileges are managed by the private cloud, who answers the file token requests from the users. The interface offered by the private cloud allows user to submit files and queries to be securely stored computed.

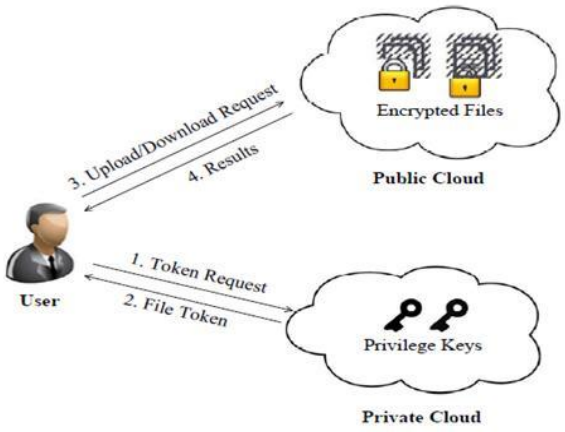

Fig. 1. Architecture for Authorized Deduplication

First if the user want to upload the files on the public cloud then user first encrypt that file with the convergent key and then sends 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 purpose of security.In the public cloud we use one algorithm for deduplication. Which is used to avoid the duplicate copies of files which is entered in the public cloud. Hence it also minimizes the bandwidth. that means we requires the less storage space for storing the files on the public cloud.In the public cloud any person that means the unauthorized person can also access or store the data so we can conclude that in the public cloud the security is not provided. In general for providing 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 uploads the many user of the public cloud because there is no security is provided in the public cloud.When user selects 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 he generated for that file. When user enter the key the private cloud checks the key for that file and if the key is correct that means user is valid then 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 same convergent key which is used at the time of encrypt that file.in this way user can make a use of thearchitecture.

VI. ADVANTAGES OF SYSTEM

1. The client is permitted to perform the duplicate copy check for records selected with the particular subject.
2. The complex subject to help stronger security by encoding the record with distinct privilege keys.

3. Decrease the storage space of the tags for reliability check. To strengthen the security of deduplication and ensure the data privacy.

VII. APPLICATION

Hybrid clouds are mainly built to suit any of the IT environment or architecture, whether it might be any enterprise wide IT network or any department. Public data which is stored can be analysed from statistical analyses which is done by social media, government entities can be used to enhance and analyse their own corporate data.

VIII. FUTURE SCOPE

It excludes the security problems that may arise in the practical deployment of the present model. Also, it increases the national security. It saves the memory by deduplicating the data and thus provides us with sufficient memory. It provides authorization to the private firms and protects the confidentiality of the important data.

IX. CONCLUSION

In this paper, the notion of authorized data deduplicationwas proposed to protect the data security byincluding differential privileges of users in the duplicatecheck. We also presented several new deduplicationconstructions supporting authorized duplicate check inhybrid cloud architecture, in which the duplicate-checktokens of files are generated by the private cloud serverwith private keys. Security analysis demonstrates thatour schemes are secure in terms of insider and outsiderattacks specified in the proposed security model. As aproof of concept, we implemented a prototype of ourproposed authorized duplicate check scheme and conducttested experiments on our prototype. We showedthat our authorized duplicate check scheme incurs minimaloverhead compared to convergent encryption andnetwork transfer.

REFERENCE

[1] OpenSSL Project. http://www.openssl.org/.

[2] P. Anderson and L. Zhang. Fast and secure laptop backups with encrypted de-duplication. In Proc. of USENIX LISA, 2010.

[3] M. Bellare, S. Keelveedhi, and T. Ristenpart. Dupless: Serveraided encryption for deduplicated storage. In USENIX Security Symposium, 2013.

[4] M. Bellare, S. Keelveedhi, and T. Ristenpart. Message-locked encryption and secure deduplication. In EUROCRYPT, pages 296–312, 2013.

[5] M. Bellare, C. Namprempre, and G. Neven. Security proofs for identity-based identification and signature schemes. J. Cryptology, 22(1):1–61, 2009.

[6] M. Bellare and A. Palacio. Gq and schnorr identification schemes: Proofs of security against impersonation under active and concurrent attacks. In CRYPTO, pages 162–177, 2002.

[7] 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.

[8] J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer.Reclaiming space from duplicate files in a serverlessdistributedfile system. In ICDCS, pages 617–624, 2002.

[9] D. Ferraiolo and R. Kuhn. Role-based access controls. In 15th NIST-NCSC National Computer Security Conf., 1992.

[10] GNU Libmicrohttpd. http://www.gnu.org/ software/libmicrohttpd/.

[11] 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.

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

[13] libcurl. http://curl.haxx.se/libcurl/.

[14] C. Ng and P. Lee. Revdedup: A reverse deduplication storage system optimized for reads to latest backups. In Proc. of APSYS, Apr 2013.

[15] W. K. Ng, Y. Wen, and H. Zhu. Private data deduplication 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.

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

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

[18] A. Rahumed, H. C. H. Chen, Y. Tang, P. P. C. Lee, and J. C. S. Lui. A secure cloud backup system with assured deletion and version control. In 3rd International Workshop on Security in Cloud Computing, 2011.

[19] R. S. Sandhu, E. J. Coyne, H. L. Feinstein, and C. E. Youman. Role-based access control models. IEEE Computer, 29:38–47, Feb 1996.

[20] J. Stanek, A. Sorniotti, E. Androulaki, and L. Kencl. A secure data deduplication scheme for cloud storage. In Technical Report, 2013.

