Volume 3, Issue 1, January-2016, pp. 1-5 ISSN (O): 2349-7084

International Journal of Computer Engineering In Research Trends

Available online at: www.ijcert.org

Secure Auditing and Deduplicating Data in Cloud

[1]K.Sudhamani, [2]P.Rama Rao, [3]R.Vara Prasad

[1]Pursuing M.Tech, CSE Branch, Dept of CSE
[2]Assistant Professor, Department of Computer Science and Engineering
[3]Assistant Professor, Department of Computer Science and Engineering
G.Pullaiah College of Engineering and Technology, Kurnool, Andhra Pradesh, India.

Abstract- The fame and extensive use of Cloud have brought huge ease for data sharing and data storage. The data sharing with a big number of participants take into account issuers like data integrity, efficiency and privacy of the owner for data. In cloud storage services one critical test is to handle rising volume of data storage in cloud. To create data management more scalable in cloud computing field, deduplication a well-known method of data compression to reduce duplicate copies of duplicate data in storage over a cloud. Even if data deduplication brings a lot of advantages in security and privacy concern occur as users' confidential data are liable to both attacks insider and outsider. A convergent encryption method imposes data privacy while making deduplication possible. Traditional deduplication systems based on convergent encryption even though offer confidentiality but do not maintain the duplicate check on basis of differential rights. This paper present, the plan of approved data deduplication planned to guard data security by counting discrepancy privileges of users in the duplicate check. Deduplication systems, users with differential privileges are added measured in duplicate check besides the data itself. To maintain stronger security the files are encrypted with differential privilege keys. Users are only permitted to carry out the copy check for files marked with the matching privileges to access. The user can confirm their occurrence of file after deduplication in cloud with the help of a third party auditor by auditing the data. Additional auditor audits and confirms the uploaded file on time. As a result, this paper generates advantages to both the storage provider and user by deduplication system and auditing method correspondingly.

Keywords— Deduplication, cloud, Cloud security, Authorized check duplicates, confidentially, auditing.

1. INTRODUCTION:

Presently cloud service provide to the users accessible high available storage and particularly parallel computing of resources at comparatively low costs. But the query is about the cloud users with different privileges store data on cloud is a most brave issue in organization cloud data storage system [7]. Deduplication is methods which make data manage more scalable in cloud computing [2]. Data deduplication describes as data compression method which eradicates second copy of repeat data in storage space. This method is use to progress storage utilization and also affect to decrease the number of bytes that must be sent before upload in data transmit. In its place to keep same satisfied data copies multiple times deduplication eliminate repetitive data and keep only one physical copy whereas submit other particular unnecessary data to that copy [3].

Deduplication can be applied to data which are in major storage, cloud storage, backup storage for replication transfers [1]. Mostly 3 types are in consideration which are as perfect deduplication process type as block level, second is file level and third is byte level by the names itself deduplicate process worked respectively on that content. Users with confidential data are worried about both outsider/insider attacks. So deduplication of data must be hold safety and privacy. But with conventional encryption dissimilar users encrypt data with their own key, which makes similar data with dissimilar user key makes different ciphertext for that data which is not capable for deduplication. The convergent encryption allows encrypt/decrypt data with convergent key on the data thus makes achievable to relate to check duplicates [3]. Therefore with uploading user's data as ciphertext to cloud determined security issues. In order to stop the un authorized access

proofs of ownership protocol can be used as privacy constraint [4]. In this Proof of ownership user can download the decrypted and acquire exact data with convergent keys by specifying its ownership. Therefore by using convergent encryption and proof of ownership both safety and privacy issues determine.

At rest the scheme can't effort on privilege level field, it means user can upload file with some set of permissions on its and on the basis on convergent encryption doesn't offer any deduplication on it [1]. As a result it will not support duplicate check with different privileges set provided by the data owner.

This paper directs to eliminate all those problems by allowing for hybrid cloud design, in which public cloud create accessibility to data owner for a given storage place which will manage by private cloud act as a proxy to allow data owner and user with security and privacy along with different permission set.

1.1. Contribution

While the data is uploaded in public cloud even if it is in encrypted form, more privacy purpose this paper is gave up by applying Public Auditing to the file uploaded in public cloud. A new Deduplication representation with the support of both security and privacy with different privilege set provided by data owner and also includes auditing. Auditing is a method in which after uploading the file by data owner an unique auditor will audit the respective file and make metadata of it's by allocating the unique audit ID number which will act as TPA.

Finally, we apply a prototype of Authorized Data deduplication as well as with auditing facility concerned in it over a hybrid network.

2. SYMMETRIC ENCRYPTION

Symmetric encryption utilizes a common secret key uses to encrypt and decrypt. A symmetric encryption includes three basic functions:
i) Key generation algorithm to generate with use of security parameter.
ii) Symmetric encryption algorithm receives secret message and then outputs cipher text.
iii) Symmetric decryption algorithm takes cipher text and then outputs original message.

2.1. Convergent encryption

Convergent encryption [3], [5] enables data privacy for deduplication procedure. A data owner or user calculates a convergent key by system and encrypts the unique data with the convergent key. User derives a tag for data copy; this tag will be used to notice duplicates. By allowing for the tag correctness property it holds the following, if two data copies are same, then their individual tags are also same. For detecting duplicates, user first sends tag to server side to confirm whether the identical copies of the data are previously present in storage or not. Convergent key and tag both are separately derived. so, the tag cannot be used to assume convergent key along with compromise data privacy. Server will store encrypted data copy and its equivalent tag. Convergent encryption system describes with four primitive functions as:
i) key generation algorithm, which maps data copy to a convergent key.
ii) Symmetric encryption algorithm, takes both the convergent key and data copy as inputs and then outputs a ciphertext
iii) Decryption algorithm, takes both the cipher text and the convergent key as inputs and then outputs the original data copy
iv) Generation algorithm for tag that maps the original data copy and outputs a tag

2.2.Proof of ownership

PoW allow user to access data which is stored in server by proving their ownership for it. PoW is interactive algorithm which runs by a prover (i.e., user) and a verifier (i.e., storage server[4]. Verifier describes a short value from a data copy.

2.3. Identification Protocol

Identification protocols have proof and verify this two phase. In first proof, a user or phase prover reveals user identity to a verifier by executing some identification proof related to his/her identity. The user input which consists of sensitive information i.e credit card number etc which the user would not share with the other users which has to maintain privacy and acts as a private key. Now the verifier performs verification with input of public

information related to the result of protocol, is that the verifier outputs either accept or reject for proof is passed or fail .

2.4. MAC-based Solution

There are two ways to make use of MAC [8]to authenticate the given data. In trifling, upload data blocks with their MACs directly to the server. Now sends the matching secret key to the TPA. After that TPA will randomly get back the blocks with their MACs and check the correctness. Apart from the high communication and computation complexities, the TPA involves the knowledge regarding the data blocks for confirmation process.

3. SYSTEM MODEL

Cloud User: A cloud user is one who needs to outsource data on public storage which acts as a public cloud in cloud computing. cloud provides authentication to the user to enter the user name and password to upload data with particular set of privileges along with that for further accessing the uploaded data to download.

Public Storage: Public Storage is an storage disk which permit to store the users data which contains authorization and not permit to upload the duplicate data. Thus save storage space and bandwidth of transmission. This uploaded data is in encrypted form, only a user with individual key can decrypt it.

Private Cloud: A private cloud acts as a proxy to allow both data owner and user to strongly perform duplicate check along with disparity permissions.

Auditor: Auditor is a TPA work as proficiency and capabilities where cloud users do not have to faith to assess the cloud storage service reliability on behalf of the user upon request.

The set of permissions and the symmetric key for each privilege is allocates and stored in private cloud. The user registers into the system, permissions are assigned to user according to identity given by the user at registration time; means on basis of situation which access by the user. The data owner with permission can upload and share a file to users, further the data owner performs identification and sends the file tag to the private server. Private cloud server checks the data owner and computes the file token and will send back the token to the data owner.

The data owner throws this file token and a request to upload a file to the storage provider. If duplicate file is found then user needs to run the PoW protocol with the storage provider to prove that user has an ownership of respective file. In the PoW result; if proof of ownership of file is approved then user will be provided a pointer for that file. And on the next case; for no duplicate is found for the file, the storage provider will be come again a signature for the result of that proof for the particular file.

To upload file user sends the privilege set as well as the proof to the private cloud server in the form of a request. The private cloud server verifies the signature first on receiving the request for the user to upload file.

Figure 1. System Model of Authorized Deduplication

Finally user computes the encryption. User encrypts the file with a key and the key is encrypted into ciphertext with each key in the file token given by the private cloud server. Then the user uploads the encrypted file, file tag, encrypted key. Assume user wants to download the file. The user first uses their key to decrypt the encrypted key and obtain key. Then the user uses to recover the original file the user may or may not be sure about the occurrence of file in the cloud. As a result, for user advantage an auditing method is used to audit the files stored in the public storage. User selects an auditor from the cloud and sends the metadata about the files going to upload in cloud to the auditor. Auditor generates audit

message or challenge to the public storage to make sure that cloud server had maintain the data file properly at the time of the audit. Public cloud storage will obtain a response message from a function of the stored data file and its Verification metadata by execution. The TPA then verifies the response through that particular users data file.

3.1 Design Goals

To develop a full fledge system three goals must be required. They are privacy preserving, security and auditing. With privacy preserving the system must possess differential authorization which is based on privilege level so that an authorized user able to get file token. If the consequence of signature verification is passed, private cloud will calculate the file token with each privileges from the privilege set given by the user, which will be returned to the user. And Authorized duplication check for a certain file will be verify by public cloud with only issued token by private on basis of private keys and privileges of the authorized user. The objective related to security of file token are unforgettable which will assure private server issued to user request and indistinguishability token doesn't give useful information to one another. To archive data confidentiality convergent encryption with higher level of privileges can be handled. To verify the user accuracy for its data auditing help us in our system.

4. AUTHORIZED DEDUPLICATION WITH AUDITING:

4.1 Main Idea: To support deduplication with authorization, the tag of a file will be determined by its privilege. For sustaining authorized access for user, a secret key will be bounded with a privilege to make a file token for it. Consider, the token is only allowed to access by user with privilege defined by the users itself. In another words, the token could only computes by the users with privilege. The token generation function could be easily denotes as a cryptographic hash function. The user with a set of privileges will assign the set of keys as Binary relations defined. If the value matches along with given two privileges, such represented as based on the background of function which include a common concept in relation of hierarchical system. More exactly, hierarchical relation is that when matches only when a higher – level privilege occurs. The target file space underlying given ciphertext is drawn from a message space of size, the public cloud server can get well after almost off-line encryptions.

We plan and implement new system which could guard security for expected message. The main idea of our method is that novel encryption key generation algorithm. To define tag generation functions and convergent keys, we will use hash functions. To carry duplicate check in traditional convergent encryption the key is derived from the file by using some cryptographic hash functions. To keep away from the deterministic key generation process, encryption key will be generated with help of private key. Encryption key can be obtained a form where all are defined as cryptographic hash functions used in system. Then the file is encrypted with another key. In this way both private cloud server and public storage cannot decrypt the ciphertext. In addition, on part of the security of symmetric encryption makes secure to the public storage. For public storage, if the file is unpredictable from, then it is protected.

5. IMPLEMENTATION:

We implemented a prototype of proposed authorized deduplication system with auditing, in which we used three model entities i.e cloud user, private cloud and public storage. *Cloud user* is a data user which performs both upload/download file on public storage. A *Private Cloud* is used as a private one which controls the private keys and handles the file token calculations. A *Public storage* wills stores and checks duplicate files present in it.We implement cryptographic process of hashing and encryption/decryption methods for storage purpose to provide cloud computing environment.

5.1 System Analysis

The security will be examined in authorization of duplicate check and confidentiality of data. For security of Duplicates check by taking into account of opponent for both internal and external, and is used to break the system by accessing cloud data or will illegal entrance to system.

5.1.1 Indistinguishiability of duplicate-check token

Protection of indistinguishability of token can also be proved based on statement of underlying message authentication code secure. Security of message for authentication code needs that adversary cannot differentiate if a code is produced

from an unknown key. In deduplication system, all privilege keys are kept secret by private cloud server. Even if a user has privilege, given a token, adversary cannot differentiate which privilege or file in the token because the user does not have knowledge of privilege key.

5.1.2 Confidentiality of Data

Convergent key in construction is not deterministic in terms of the file. Depend on privilege secret key stored by private cloud server and unknown to the opponent. Thus, if opponent does not join together with the private cloud server, privacy of our second construction is semantically secure for both conventional and unpredictable file. If not in case of, they join together then confidentiality of file will be reduced to convergent encryption because encryption key is deterministic.

Auditor achievement deals with appropriate handling the data content of the users with exact generated key.

6. CONCLUSION

The fame and extensive use of Cloud have brought great handiness for data sharing and collection. One vital challenge of cloud storage services is to handle the ever-increasing volume of data content stored. To make data managed scalable in cloud computing methods, Data deduplication has been a well-known method presented here. Data deduplication is a specific data compression technique for eliminating duplicate copies of repeating data in storage. Although data deduplication contributes a lot of advantages, along with security and privacy concerns arise as user's confidential data are liable to both inside and outside attacks. In this paper, the design of authorized data deduplication was projected to protect data security by counting differential privileges of users in the duplicate check. Contrast from traditional deduplication systems, the discrepancy privileges sets of users are further considered in duplicate check. . For support of stronger security the files are encrypted with differential privilege keys. The user is only allowed to carry out the duplicate check for files noticeable with the corresponding privileges. The user can verify their presence of file after deduplication in cloud by auditing the data with the help of a third party auditor. The auditor audits and validates the uploaded file on time. As a result, the paper presents profits to both the storage provider and user by deduplication method and auditing method correspondingly.

REFERENCES

1. Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou Jin Li, "A Hybrid Cloud Approach for Secure Authorized," *IEEE Transactions on Parallel and Distributed Systems*, vol. pp, pp. 1-12, 2014.
2. S.Quinlan and S. Dorward., "Venti: a new approach to archival storage," *USENIX FAST*, Jan 2002.
3. A. Adya, W. J. Bolosky, D. Simon, and M. Theimer. J. R. Douceur, "Reclaiming space from duplicate files in a serverless distributed," *ICDCS*, pp. 617-624, 2002.
4. D. Harnik, B. Pinkas, and A. Shulman-Peleg. S. Halevi, "Proofs of ownership in remote storage systems.," *ACM Conference on Computer and Communications Security*, pp. 491-500, 2011.
5. Sriram Keelveedhi, Thomas Ristenpart Mihir Bellare, "Message-locked encryption and secure deduplication," in *Springer Berlin Heidelberg,International Association for Cryptologic Research*, Advances in Cryptology – EUROCRYPT 2013, Athens, Greece, March 2013, pp. 296-312.
6. S. Nurnberger, A. Sadeghi, and T. Schneider. S. Bugiel, "Twin clouds: An architecture for secure cloud computing.," *Workshop on Cryptography and Security in Clouds (WCSC 2011)*, 2011.
7. Edward J. Coynek, Hal L. Feinsteink and Charles E. Youmank Ravi S. Sandhu, "Role-Based Access Control Models," *IEEE Computer*, vol. 29, pp. 38-47, Feb 1996.
8. Sherman S.-M. Chow, Qian Wang, Kui Ren, and Wenjing Lou Cong Wang, "Privacy-Preserving Public Auditing for Secure Cloud Storage," *Computers, IEEE Transactions*, vol. 62, no. 2, pp. 362 - 375, Feb 2013.
9. Chanathip Namprempre, Gregory Neven Mihir Bellare, "Security Proofs for Identity-Based Identification and Signature Schemes," *Journal of Cryptology,Springer-Verlag*, vol. 22, no. 1, pp. 1-61, January 2009.

