International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 04 Issue 01
January2017

A Survey on Secure Auditing and Deduplication Data in Cloud

[1]N.Srinivas & [2]Mrs.T.Swathi
[1]M.Tech (CSE) from JAGRUTI INSTITUTE OF ENGINEERING AND TECHNOLOGY
[2] Assistant Professor, Department of Computer Science and Engineering, JAGRUTI INSTITUTE OF ENGINEERING AND TECHNOLOGY, Telangana State, India.

ABSRACT: *As the cloud computing technology develops throughout the last decade, outsourcing data to the cloud server for storage becomes a attractive development, that advantages in economical efforts on significant information maintenance and management. Nevertheless, since the outsourced cloud storage isn't totally trustworthy, it raises security issues on the way to understand data deduplication in cloud whereas achieving integrity auditing. during this work, we have a tendency to study the matter of integrity auditing and secure deduplication on cloud information. Specifically, aiming at achieving each information integrity and deduplication in cloud, we have a tendency to propose 2 secure systems, particularly SecCloud and SecCloud+. SecCloud introduces associate auditing entity with a maintenance of a MapReduce cloud, that helps shoppers generate information tags before uploading additionally as audit the integrity of knowledge having been keep in cloud. Compared with previous work, the computation by user in SecCloud is greatly reduced throughout the file uploading and auditing phases. SecCloud+ is meant impelled by the actual fact that customers continually need to encode their information before uploading, and allows integrity auditing and secure deduplication on encrypted information.*

Keywords: Deduplication, Block Level Check, authorized duplicate check, confidentiality, File level Check, Metadata Supervisor.

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.

2. RELATED WORK

1**) A view of cloud computing** AUTHORS: M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia Cloud computing, the long-held dream of computing as a utility, has the potential to transform a large part of the IT industry, making software even more attractive as a service and shaping the way IT hardware is designed and purchased. Developers with innovative ideas for new Internet services no longer require the large capital outlays in hardware to deploy their service or the human expense to operate it. They need not be concerned about overprovisioning for a service whose popularity does not meet their predictions, thus wasting costly resources, or underprovisioning for one that becomes wildly popular, thus missing potential customers and revenue. Moreover, companies with large batchoriented tasks can get results as quickly as their programs can scale, since using 1,000 servers for one hour costs no more than using one server for 1,000 hours. This elasticity of resources, without paying a premium for large scale, is unprecedented in the history of IT.

2) **Secure and constant cost public cloud storage auditing with deduplication** AUTHORS: J. Yuan and S. Yu Data integrity and storage efficiency are two important requirements for cloud storage. Proof of Retrievability (POR) and Proof of Data Possession (PDP) techniques assure data integrity for cloud storage. Proof of Ownership (POW) improves storage efficiency by securely removing unnecessarily duplicated data on the storage server. However, trivial combination of the two techniques, in order to achieve both data integrity and storage efficiency, results in nontrivial duplication of metadata (i.e., authentication tags), which contradicts the objectives of POW. Recent attempts to this problem introduce tremendous computational and communication costs and have also been proven not secure.

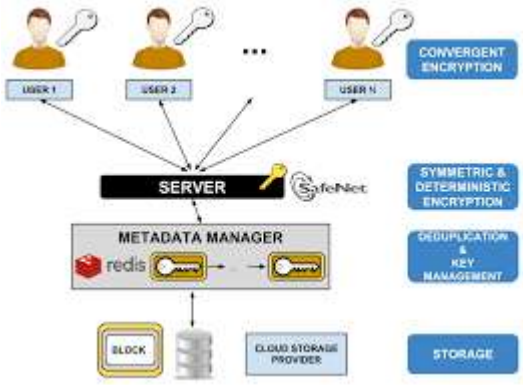
3. SYSTEM MODEL

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



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.

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.

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.

REFERECES:

[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, "RoleBased 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, "PrivacyPreserving 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 IdentityBased Identification and Signature Schemes," Journal of Cryptology,Springer-Verlag, vol. 22, no. 1, pp. 1-61, January 2009.

[10] H. Wang, "Proxy provable data possession in public clouds," *IEEE Transactions on Services Computing*, vol. 6, no. 4, pp. 551–559, 2013.

[11] Y. Zhu, H. Hu, G.-J. Ahn, and M. Yu, "Cooperative provable data possession for integrity verification in multicloud storage," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 12, pp. 2231– 2244, 2012.

[12] H. Shacham and B. Waters, "Compact proofs of retrievability," in *Proceedings of the 14th International Conference on the Theory and Application of Cryptology and Information Security: Advances in Cryptology*, ser. ASIACRYPT '08. Springer Berlin Heidelberg, 2008, pp. 90–107.

[13] Q. Wang, C. Wang, J. Li, K. Ren, and W. Lou, "Enabling public verifiability and data dynamics for storage security in cloud computing," in *Computer Security – ESORICS 2009*, M. Backes and P. Ning, Eds., vol. 5789. Springer Berlin Heidelberg, 2009, pp. 355–370.

[14] J. Xu and E.-C. Chang, "Towards efficient proofs of retrievability," in *Proceedings of the 7th ACM Symposium on Information, Computer and Communications Security*, ser. ASIACCS '12. New York, NY, USA: ACM, 2012, pp. 79–80.