International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 04 Issue 06
May 2017

A Study on Deduplication and Secure Public Auditing In Cloud Storage

NAME: UTKARSH GUPTA
M.TECH CSE
GD GOENKA UNIVERSITY
GURUGRAM

ABSTRACT: Storing sensitive data to cloud storage becomes an important trend, the data owner's getting burden free which benefits in sparing efforts on heavy data maintenance and management. The Data which is outsourced to cloud is not secure why because the cloud is Untrusted, and it leads privacy concerns on how to assure data deduplication in cloud while getting integrity auditing. In this paper , we study the problem of integrity auditing and secure deduplication on cloud data. Specifically, aiming at getting both data integrity and deduplication in cloud, we present two secure systems, namely SecCloud and SecCloud+. SecCloud introduces an auditing entity with a maintenance of a MapReduce cloud, which helps clients create data tags before uploading as well as audit the integrity of data having been saved in cloud. Compared with previous work, the computation by user in SecCloud is greatly reduced during the file uploading and auditing phases. SecCloud+ is designed motivated by the fact that customers always want to encrypt their data before uploading, and enables integrity auditing and secure deduplication on encrypted data.

Key Words: Cloud Storage, Data deduplicating, Secure auditing.

INTRODUCTION:

Cloud computing offers service as a resource ,cloud provides IaaS, PaaS,SaaS etc cloud has advanced computational Resource power and which provided data sharing and data storing facility. Cloud computing is a shared pool of configurable computing resources, on-demand network access and provisioned by the service provider [1].It is cost saving but other hand it has major concern of security. Cloud storage is one of the attractive trend which provide benefits to the customer like cost saving, mobility and scalable service. Cloud Data Storage Service A cloud data storage service involves 3 main entities. i. CSP – Control over file insertion, file access, file deletion and at the time of user presents in the network and trying to access the cloud data's. ii. Third Party Auditor (TPA) checks – TPA check on the correctness of cloud data. iii. Users Access – Availability of the cloud data as per demand services. Data security is a major concern related to cloud computing. Most of the cloud service provider provides some facility for security and privacy which mainly includes 4 types of data items

i) usage data ii) sensitive data iii) Personally identifiable information iv) Uniques device identities

[B] Data integrity Integrity is another name consistency. It is a major factor which affects the cloud performance. Data integrity has a protocol for writing of the data in a reliable manner to the persistent data storages which when retrieved is in the same format without any changes. Maintaining integrity of shared data is difficult task. There are number of mechanisms have been proposed [2], [3], [4], [5], [6], & [7] to maintain integrity of data. Integrity is most

important of all the security issues in cloud data storages as it ensure completeness of data as well as that the available data is correct, easily accessible to authorized user, consistent and of high quality. There are three types of integrity constraints: Domain integrity, Referential integrity, Entity integrity. The correctness of data storage and computation compromised on the cloud due to the less of the control of data owners on data. Secure cloud computing always focuses on the cloud data storage security and cloud computing security. Safety of the data stored on the cloud has been compromised in many cases for monetary profit. To avoid this it is essential to maintain security and privacy of data and cloud computing by using different techniques and mechanisms. [C] Deduplication Data deduplication is one of the special techniques which used as a data compression technique, which helps in eliminating duplicate copies of repetitive data, here cloud server stores single copy of data file. This technique is provide benefits like saving network bandwidth but on other hand in hybrid cloud which is combination of public and private cloud deduplication may lead to loss of sensitive information. Deduplication technique has two categories, which are based data units – 1. File Level Deduplication - Here file is considered as a one data unit, hash value of file is used as its identifier. During deduplication check if two or more files have similar has value, then they consider that files with same contents and only one copy will be stored. 2. Block Level Deduplication - Here file is divided into small data blocks, these data blocks are fixed-size or variable size, to check deduplication hash value is computed on each data block.

EXISTING SYSTEM:

A number of deduplication systems have been proposed based on various deduplication strategies. such as clientside or server-side deduplications , file-level or block-level deduplication. Bellare et al formalized this primitive as message-locked

encryption, and explored its application in space efficient secure outsourced storage. Li addressed the key-management issue in block-level

deduplication by distributing these keys across multiple servers after encrypting the files. Bellare et al showed how to protect data confidentiality by

transforming the predictable message into unpredictable message. The first problem is integrity auditing. The cloud server is

able to relieve clients from the heavy burden of storage management and maintenance. The most difference of cloud storage from traditional in-house storage is that the data is transferred via Internet and stored in an uncertain domain ,not under control of the clients at all, which inevitably raises clients great concerns on the integrity of their data. The second problem is secure deduplication. The rapid

adoption of cloud services is accompanied by increasing volumes of data stored at remote cloud servers. Among these remote stored files, most of them are duplicated: according to a recent survey by EMC, 75% of recent digital data is duplicated copies.

Unfortunately, this action of deduplication would lead to a number of threats potentially affecting the storage system, for example, a server telling a client that it (i.e., the client) does not need to send the file reveals that some other client has the exact same file, which could be sensitive sometimes. These

attacks originate from the reason that the proof that the client owns a given file (or block of data) is solely based on static, short value (in most cases the hash of the file).

PROPOSED SYSTEM:

In this paper, It has shown how to design secure deduplication systems with higher reliability in cloud computing. By introducing the distributed cloud storage servers into deduplication systems to provide better fault tolerance. To further protect data confidentiality, the secret sharing technique is utilized, which is also compatible with the distributed storage systems. In more details, a file is first split and encoded into fragments by using the technique of secret sharing, instead of encryption mechanisms. These shares will be distributed across multiple independent storage servers. Furthermore, to support deduplication, a short cryptographic hash value of the content will also be computed and sent to each storage server as the fingerprint of the fragment stored at each server. Only the data owner who first uploads the data is required to compute and distribute such secret shares, while all following users who own the same data copy do not need to compute and store these shares any more. To recover data copies, users must access a minimum number of storage servers through authentication and obtain the secret shares to reconstruct the data. In other words, the secret shares of data will only be accessible by the authorized users who own the corresponding data copy. Four new secure deduplication systems are proposed to provide efficient deduplication with high reliability for filelevel and block-level deduplication, respectively. The secret splitting technique, instead of traditional encryption methods, is utilized to protect data confidentiality. Specifically, data are split into fragments by using secure secret sharing schemes and stored at different servers.

CONCLUSION:

Providing both data integrity and deduplication in cloud, we present SecCloud and SecCloud+. SecCloud proposes an auditing entity with maintenance of a MapReduce cloud, which helps clients create data tags before uploading as well as audit the integrity of data having been stored in cloud. In addition, SecCoud enables secure deduplication through ipresenting a Proof of Ownership protocol and avoiding the leakage of side channel information in data deduplication. Compared with previous work, the computation by user in SecCloud is greatly decreased during the file uploading and auditing phases. SecCloud+ is an advanced construction motivated by the fact that customers always want to encrypt their data before uploading, and allows for integrity auditing and secure deduplication directly on encrypted data.

REFERENCES:

[1] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski,G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "A view of cloud computing," Communication of the ACM, vol. 53, no. 4, pp.50–58, 2010.

[2] J. Yuan and S. Yu, "Secure and constant cost public cloud storage auditing with deduplication," in IEEE Conference on Communications and Network Security (CNS), 2013, pp. 145–153.

[3] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg, "Proofs of ownership in

remote storage systems," in Proceedings of the 18th ACM Conference on Computer and Communications Security. ACM, 2011,pp. 491–500.

[4] K. Kiran Kumar, K. Padmaja, P. Radha Krishna, "Automatic Protocol Blocker for Privacy-Preserving Public Auditing in Cloud Computing", International Journal of Computer science and Technology, vol. 3 pp, ISSN. 0976-8491(Online), pp. 936-940, ISSN: 2229-4333 (Print), March 2012.

[5] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson,and D. Song, "Provable Data Possession at Untrusted Stores,"in the Proceedings of ACM CCS 2007, pp. 598–610.

[6] G. Ateniese, R. Burns, R. Curtmola, J. Herring, O. Khan, L. Kissner, Z. Peterson, and D. Song, "Remote data checking using provable data possession," ACM Trans. Inf. Syst. Secur., vol. 14, no. 1, pp. 12:1–12:34, 2011.

[7] G. Ateniese, R. Di Pietro, L. V. Mancini, and G. Tsudik, "Scalable and efficient provable data possession," in Proceedings of the 4th International Conference on Security and Privacy in Communication Netowrks, ser. SecureComm '08. New York, NY, USA: ACM, 2008, pp. 9:1– 9:10.

[8] C. Erway, A. K¨upc¸ ¨u, C. Papamanthou, and R. Tamassia, "Dynamic provable data possession," in Proceedings of the 16th ACM Conference on Computer and Communications Security, ser. CCS '09. New York, NY, USA: ACM, 2009, pp. 213–222.

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

[10] 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.

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

[12] M. W. Storer, K. Greenan, D. D. E. Long, and E. L. Miller. Secure data deduplication. In Proc. of StorageSS, 2008

[13] M. Shyamala Devi, V.Vimal Khanna, Naveen Balaji"Enhanced Dynamic Whole File De-Duplication (DWFD) for Space Optimization in Private Cloud Storage Backup", IACSIT, August, 2014.

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

[15] E. Stefanov, M. van Dijk, A. Juels, and A. Oprea, "Iris: A scalable cloud file system with efficient integrity checks," in Proceedings of the 28th Annual Computer Security Applications Conference, ser. ACSAC '12. New York, NY, USA: ACM, 2012, pp. 229–238.

[16] M. Azraoui, K. Elkhiyaoui, R. Molva, and M. O¨ nen, "Stealthguard: Proofs of retrievability with hidden watchdogs," in Computer Security - ESORICS 2014, ser. Lecture Notes in Computer Science, M. Kutyłowski and J. Vaidya, Eds., vol. 8712. Springer International Publishing, 2014, pp. 239–256.

[17] J. Li, X. Tan, X. Chen, and D. Wong, "An efficient proof of retrievability with public auditing in cloud computing," in 5th International Conference on Intelligent Networking and Collaborative Systems (INCoS), 2013, pp. 93–98.