ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2016

Efficient & Secure Auditing and Data Deduplication in Cloud Computing

Waghule Ashwini[1], Purbiya Divya[2], Goski Soundarya[3], Nalkar Rutuja[4], Prof. Aher S. M[5]

B.E Student, Dept. of Computer Engineering, SCSMCOE, Nepti, Ahmednagar, Maharashtra, India[1,2,3,4]

Asst. Professor, Dept. of Computer Engineering, SCSMCOE, Nepti, Ahmednagar, Maharashtra, India[5]

ABSTRACT: Data deduplication is a technique for reducing the amount of storage space an organizationneeds to save its data. In most organizations, the storage systems contain duplicatecopies of many pieces of data. For example, the same file may be saved in several different places by different users, or two or more less that aren't identical may still include muchof the same data. Deduplication eliminates these extra copies by saving just one copy ofthe data and replacing the other copies with pointers that lead back to the original copy.Companies frequently use deduplication in backup and disaster recovery applications,but it can be used to free up space in primary storage as well. To avoid this duplication of data and to maintain the confidentiality in the cloud we using the concept of Hybrid cloud. To protect the confidentiality of sensitive data while supporting deduplication,the convergent encryption technique has been proposed to encrypt the data before outsourcing.To better protect data security, this paper makes the first attempt to formallyaddress the problem of authorized data deduplication.In this work, we study the problem of integrity auditingand secure deduplication on cloud data. Specifically, aiming atachieving both data integrity and deduplication in cloud,

KEYWORDS: Seccloud , seccloud+, integrity auditing, secure de-duplication , proof of ownership convergent encryption.

I. INTRODUCTION

Even though cloud storage system has been mostly adopted, it fails to accommodate some important emerging needs such as the capability of auditing integrity of c loud files by cloud clients and detecting duplicated files by c loud servers. We disclose both problems below. The first problem is integrity auditing. The cloud server is able to relieve clients from the bulky 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. These concerns originate from the fact that the cloud storage is affected to security threats from both outside and inside of the cloud, and the uncontrolled cloud servers may passively hide some data loss incidents from the clients to maintain their reputation. What is more serious is that for saving money and space, the cloud servers might eve n actively and deliberately discard barely accessed data files belonging to an ordinary client. Considering the large size of the outsourced data files and the clients' constrained resource capabilities, the first problem is generalized as how can the client efficiently perform regularly integrity verifications even without the local copy of data file.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. Deduplication is methods which make data manage more scalable in cloud computing . 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 affectto decreasethe number of bytes that must be sent before upload in data transmit. In its placeto keep same satisfied data copies multiple times deduplication eliminaterepetitive data and keep only one physical copy whereassubmitother particularunnecessarydata to that copy Deduplication can be applied to data which are in major storage, cloud storage, backup storage for replication transfers. Mostly 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 it self 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

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2016

dissimilar users encryptd ata 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.

II. LITERATURE SURVEY

Sr no.	Author	Title	Technique used
1	Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou	"A Hybrid Cloud Approach for Secure Authorized Deduplication.."	Deduplication of data or file is based on only file name. block level checking is not perform in this technique.
2	Jin Li, Xiaofeng Chen, Mingqiang Li, Jingwei Li, Patrick P.C. Lee, and Wenjing Lou	"Secure Deduplication with Efficient and Reliable Convergent Key Management"	Each user holds independent master key for encrypting all file which are uploaded by same user.
3	Peter Christen.	"A Survey of Indexing Techniques for Scalable Record Linkage and Deduplication."	In database there may be multiple record entries of duplicate information .so that duplicate get reduced using indexing technique to avoid duplication of record in database.
4	Mihir Bellare, Sriram Keelveedhi, Thomas Ristenpart	"DupLESS: Server-Aided Encryption for Deduplicated Storage"	Author proposed architecture that pro-vides secure deduplicated storage resisting brute-force attacks, and realize it in a system called DupLESS.

Table 1. Deduplication Methods

III. PROPOSED ALGORITHM

SecCloud introduces an auditing entity with a maintenance of a MapReduce cloud, which helps clients generate data tags before uploading as well as audit the integrity of data having been stored in cloud. This design fixes the issue of previous work that the computational load at user or auditor is too huge for tag generation. For completeness of fine-grained, the functionality of auditing designed in SecCoud is supported on both block level and sector level.

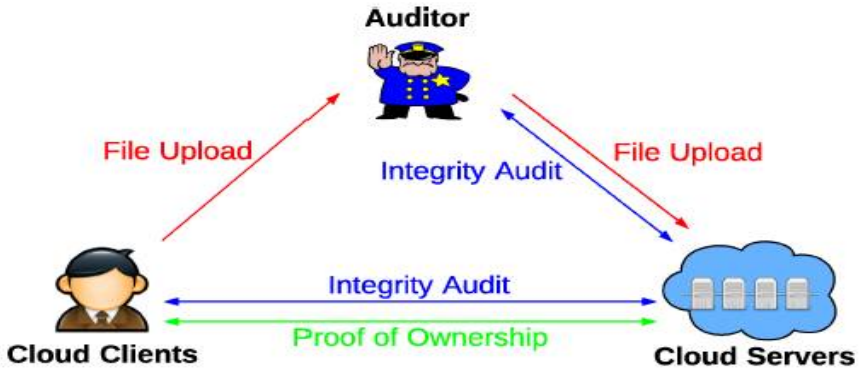

Fig. 2 System Architecture

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2016

In addition, SecCoud also enables secure deduplication. Notice that the "security" considered in SecCoud is the prevention of leakage of side channel information. In order to prevent the leakage of such side channel information, the tradition of and design a proof of ownership protocol between clients and cloud servers, which allows clients to prove to cloud servers that they exactly own the target data. Motivated by the fact that customers always want to encrypt their data before uploading, for reasons ranging from personal SecCloud as with [4] and propose the SecCloud+ schema. Besides supporting integrity auditing and secure deduplication, SecCloud+ enables the guarantee of file confidentiality. Specifically, thanks to the property of deterministic encryption in convergent encryption, we propose a method of directly auditing integrity on encrypted data. The challenge of deduplication on encrypted is the prevention of dictionary attack. As with, we make a modification on convergent encryption such that the convergent key of file is generated and controlled by a secret "seed", such that any adversary could not directly derive the convergent key from the content of file and the dictionary attack is prevented.

SECCLOUD:

- Integrity Auditing. The first design goal of this work isto provide the capability of verifying correctness of theremotely stored data.
- Secure Deduplication. The second design goal of thiswork is secure deduplication. In other words, it requiresthat the cloud server is able to reduce the storage spaceby keeping only one copy of the same file.
- Cost-Effective. The computational overhead for providingintegrity auditing and secure deduplication shouldnot represent a major additional cost to traditional cloud storage.

SECCLOUD+:

- File Confidentiality. The design goal of file confidentialityrequires to prevent the cloud servers fromaccessing the content of files.

convergent encryption scheme canbe defined with four primitive functions:

• KeyGen(F) : The key generation algorithm takes a filecontent F as input and outputs the convergent key ckFof F;

• Encrypt(ckF;F) : The encryption algorithm takes theconvergent key ckF and file content F as input andoutputs the ciphertext ctF;

• Decrypt(ckF; ctF) : The decryption algorithm takes theconvergent key ckF and ciphertext ctF as input andoutputs the plain file F;

• TagGen(F) : The tag generation algorithm takes a filecontent F as input and outputs the tag tagF of F. Noticethat in this paper, we also allow TagGen(·) to generatethe (same) tag from the corresponding ciphertext.

IV. CONCLUSION AND FUTURE WORK

We are focusing to achieve both data integrity and data de-duplication in cloud, we propose SecCloud andSecCloud+. SecCloud introduce an auditing entity with maintenance of a MapReduce cloud, which helps the clientsto generate data tags before uploading and audit the integrity of data having been stored in cloud. In addit ion,SecCoud enables secure de-duplication by introducing a Proof of Ownership protocol (POP) and preventing theleakage of side channel informat ion in de-duplication. We compared with the existing work, the computation byuser in SecCloud is greatly reduced during the file uploading and auditing phases. Sec - Cloud+ is an advancedconstruction motivated by the fact that customers always want their data to be encrypted before uploading, andallows for integrity auditing as well as secure de-duplication 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 viewof 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 storageauditing with deduplication," in IEEE Conference on Communicationsand Network Security (CNS), 2013, pp. 145–153.

[3] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg, "Proofs ofownership in remote storage systems," in Proceedings of the 18th ACM Conference on Computer and Communications Security. ACM, 2011,pp. 491–500.

[4] S. Keelveedhi, M. Bellare, and T. Ristenpart, "Dupless: Serveraidedencryption for deduplicated storage," in Proceedings of the22Nd USENIX Conference on Security, ser. SEC'13. Washington,D.C.: USENIX Association, 2013, pp. 179–194.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 9, September 2016

[5] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson,and D. Song, "Provable data possession at untrusted stores," in Proceedingsof the 14th ACM Conference on Computer and CommunicationsSecurity, ser. CCS '07. New York, NY, USA: ACM, 2007.
[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, "Scalableand efficient provable data possession," in Proceedings of the 4thInternational Conference on Security and Privacy in CommunicationNetowrks, ser. SecureComm '08. New York, NY, USA: ACM, 2008.
[8] C. Erway, A. K¨upc¸ ¨u, C. Papamanthou, and R. Tamassia, "Dynamicprovable 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] F. Seb´e, J. Domingo-Ferrer, A. Martinez-Balleste, Y. Deswarte, andJ.-J. Quisquater, "Efficient remote data possession checking in critical Information infrastructures," IEEE Trans. on Knowl. and Data Eng. vol. 20, no. 8, pp. 1034–1038, 2008.
[10] Q. Wang, C. Wang, J. Li, K. Ren, and W. Lou, "Enabling publicverifiability 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.
[11] J. Xu and E.-C. Chang, "Towards efficient proofs of retrievability," inProceedings of the 7th ACM Symposium on Information, Computer and Communications Security, ser. ASIACCS '12. New York, NY, USA:ACM, 2012, pp. 79–80.