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 6, June 2016

Implementation of Secure Distributed Deduplication and Auditing Systems with Improved Reliability

Kalyani Zodge, Amruta Amune

PG Student, Department of Computer Engineering, G.H.Raisoni College of Engineering and Management Chas, Ahmednagar, India

Assistant Professor, Department of Computer Engineering, G.H.Raisoni College of Engineering and Management Chas, Ahmednagar, India

ABSTRACT: Today we all know how fast data is being generated & how this data is important to different organisations &enterprises. Hence organisations all around globe are looking forward to store this data effectively to avoid data loss & also store all in compressed format & secure manner. Data deduplication is emerging trend which give solution to store more information in less space,as provide benefits like less maintainence of storage infrastructure & cost efficiency. As there is only one copy for each file stored in cloud even that file owned by large no of users. Therefore deduplication improves storage utilization & affect reliability. The challenge of privacy for sensitive data also arises when that data outsourced by user to cloud.To overcome above security challenge this paper present system that provides novel distributed deduplication system with higher reliability by distributing the data chunks across multiple cloud servers. Data confidentiality achived by implementing secret sharing schemes. Additionally this paper introduces (TPA)to enable public auditability for cloud storage so that user can use TPA to check integrity of outsourced data & be worry free, we extend that enable TPA to perform audits for multiple users efficiently and simultaneously that is batch auditing. Furthermore this paper introduces a KDC authority that is used for key management..

KEYWORDS: Deduplication,secretssharing, reliability, auditing

I. INTRODUCTION

As there is explosive growth of digital data, data deduplication techniques are used to backup data and reduce network and storage overhead by removing redundancy among data. Instead of keeping multiple copies with the similar content, deduplication eliminates redundant data by keeping only single physical copy and give reference to other redundant data to that copy. Deduplication has become popular in both field acdemics & industry because it improves storage utilization and save storage space.

Specially, with the advent of cloud storage, data deduplication procedure grow to be more gorgeous and essential for the management of ever-increasing quantity of data in cloud storage services which inspires Endeavour and club to outsource data storage to third-party cloud providers, If we consider some of the examples as proofs: Today's cloud storage services, such as, Google Drive, Drop box have been pertaining deduplication to save the network bandwidth and the storage cost.There are two types of deduplication with respect to size: (I) file-level deduplication:which discovers redundanciesbetween different files and elliminates these redundancies to reduce storage space. (ii) blockleveldeduplication:which discovers and eliminate redundanciesbetween data blocks**.**

To give ensurity of the data integrity and save the cloud users'computation resources as well as online burden, it is ofimportance to enable public auditing service for cloud data storage, so that users may resort to an independent third-party auditor (TPA) to audit the outsourced data when needed. The TPA is nothing but who has expertise and capabilities that users do not have, can periodically check the integrity of all the data stored in the cloud on behalf of the users, which gives easiest way for the users to give ensurity of their storage correctness in the cloud. In a word, enabling public auditing services will play an important role for this nascent cloud economy to become fully

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 6, June 2016

established, where users will need ways to assess risk and gain trust in the cloud.In a word, enabling public auditing services will take part in an important role for this nascent cloud economy to become fully reputable; where users will requires ways to assess risk and gain trust in the cloud.

The rest of the paper is arranged as follows. In section II the related work is discussed. In section III, we propose the system model for our deduplication system. We support the discussion by considering problem definition, barriers ofexisting systems, mathematical model. In the next section the experimental results have been discussed and then we conclude paper in section .

II. RELATED WORK

Reliable Deduplication schemes: Data deduplication methods are interesting systems that are extensively working for data backup in enterprise atmospheres to reduce network as well as storage overhead by detecting &rejecting redundancy between data blocks. Li et al. [11] introduced that how to achieve reliable key management indeduplication system. Though, they did not remark about the application of the reliable deduplication for encrypted files or block of files. Li et al[2] addressed the key-management problem in block-level deduplication through distributing these keys across the multiple servers after encrypting the files. Li, et al. [3], presented the convergent dispersal that supports keyless security as well as deduplication aimed at cloud-of-clouds storage, for the creation of reliable deduplication meant for user files. On the other hand, all of these works have not measured and achieved the tag consistency as well as integrity in the production.Bellare et al. [4] dignified this primitive such as Message-Locked Encryption (MLE), as well as explored its application in space efficient secure outsourced storage. They provide definitions together for privacy as well as for a form of integrity that can call as tag consistency. They make connections through deterministic encryption, the hash functions secure on correlated inputs and the sample-then-extract model to deliver systems under dissimilar assumptions and for different types of message source.Harnik et al. [5] presented various numbers of attacks that can cause data leakage/loss in a cloud storage application based on client-side deduplication. They learn the privacy implications of cross-user deduplication. They prove how deduplication can be considered as a side channel that can reveals information regarding the contents of files of other users. In a different situation, deduplication can be considered as a covert channel by which malicious software can communicate with its control center, regardless of any firewall settings at the attacked machine.
Ateniese et al. [6] are the primary to believe public auditability in their “provable data possession” (PDP) model for ensuring tenure of data files on untrusted storages. They use the RSA-based homomorphic linear authenticators for auditing outsourced data and propose arbitrarily sampling a few blocks of the file. Among their two projected schemes, the one with public auditability depiction the linear arrangement of sampled blocks to external auditor. When used directly, their protocol is not conclusively privacy preserving, and thus may leak user data information to the external auditor. Shah et al. [7] suggest introducing a TPA to keep online storage honest by making first encryption of data then distributing a number of precompiled symmetric-keyed hash value over the encrypted data to the auditor. The auditor authenticates the data file integrity and the server’s control of a beforehand committed decryption key. This system only works for encrypted files, necessitates the auditor to maintain state, and suffers from bordered usage, which potentially brings in online load to users when the keyed hashes are used up.

III PROPOSED METHODOLOGY AND DESIGN

The proposed system is presented for carrying out secured deduplication and auditing process. It has many significant features.This section will describe the system design, the entities, consider while developing the system and the implementation plan.

A.Problem Definition:

To increase the amount of information that can be stored on cloud storage provider by saving bandwidth and to eliminate redundant copies of data to preserve confidentiality of sensitive data while supporting deduplication & auditing makes privacy preservation that is maintain data integrity.

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 6, June 2016

B.System overview:

In this paper, it shows how to design secure deduplication techniques with higher reliability in the cloud computing. This paper introduces the distributed cloud storage servers into deduplication methods to provide improved fault tolerance. To additional protect the data confidentiality; the secret sharing method is utilized that also compatible with distributed storage approaches. A file is previously split and encoded into fragments by using the system of secret sharing (SS), instead of encryption mechanisms. These secret shares will be distributed through multiple independent storage cloud servers. Additionally, to useful deduplication, a short cryptographic hash value of the content will also be computed and sent to each storage server as fingerprint of the fragment keep at every server. only the information owner who initial uploads the information is needed to calculate and distribute such secret shares, whereas all following users who own an equivalent information copy do not need to calculate and store these shares any more. To recover knowledge copies, users should access a minimum number of storage servers through authentication and obtain the secret shares to reconstruct the information. In different words, the secret shares information can only be accessible by the authorized users who own the corresponding data copy. Four new secure deduplication systems are projected to provide efficient deduplication with high reliability for file-level and block-level deduplication. The secret splitting technique, rather than traditional encryption strategies, is utilized to protect information confidentiality. Specifically, information are split into no of fragments by using secure secret sharing schemes and stored at totally different servers. The projected constructions support each file-level and block-level deduplication. Security analysis demonstrates that the projected deduplication systems are secure in terms of the definitions specified in the projected security model. In additional details, confidentiality, reliability and integrity can be achieved in proposed system. Two types of collusion attacks are considered in solutions. Above all, the information remains secure even if the adversary controls a restricted number of storage servers. The deduplication systems implemented exploiting the Ramp secret sharing scheme that allows high responsibility and confidentiality levels. The analysis results demonstrate that the new projected constructions are efficient and therefore the redundancies are optimized and comparable the other storage system supporting identical level of reliability.

The proposed system motivate the general public auditing system of information storage security in cloud computing and provide a privacy-preservation auditing protocol. To the simplest of our information, our scheme is that to support efficient and scalable privacy-preserving public storage auditing in cloud. Specifically, our scheme achieves batch auditing wherever multiple auditing tasks from different users will be performed at the same time by the TPA in a privacy-preserving manner. The assumption information integrity threats toward users' data are that they will come from each internal and external attack at Cloud Server (CS). These can be hardware failures, software bugs, bugs in the network path, economically motivated hackers, malicious or accidental management errors, etc. Besides, CS will be self-interested. The architecture consist KDC for key management. The authentication as well as access control are both collusion resistant, means that no two users can collude and access data or authenticate themselves, if they are individually not authorized.

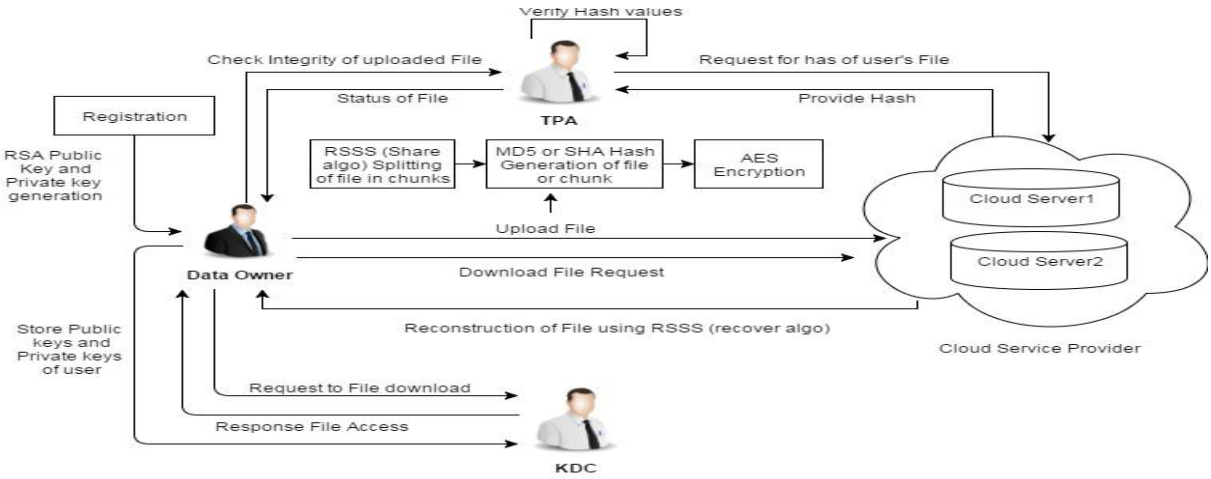

Figure.1 Proposed System Architecture.

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 6, June 2016

C Entities :

1. Data Owner

The user requires outsourcing data storage to the CSP and access data later. In a storage system underneath deduplication, the user only uploads unique data but does not upload other duplicate data to save the upload bandwidth. Furthermore, the fault tolerance is required by users in the system to provide higher reliability.

2. Key Distributer Center

It is an indispensable entity which is responsible for generating, distributing and managing all the private keys, and is faithful by all the other entities contributed in the system.

3. TPA

The TPA subjects an audit message or protest to the cloud server to make sure that the cloud server has retained the data file F properly at the time of the audit. The (CSP) cloud server will derive a response message by executing Reproof using F and its authentication metadata as inputs. The TPA then verifies the reply via Verify Proof.

4. CSP

The CSP is an entity that provides the outsourcing data storage service for the users. In the deduplication system, when users own and store the same content, the CSP will only store a single copy of these files and retain only exclusive data.

C Proposed Methodology :

1. User Registration Stage

i. *Key Generation Algorithm:*

When user wants to upload file on cloud user need to first register and that time private key ,public key generated for that user by using RSA.These keys are passed to KDC for storage purpose.

- RSA Algorithm

  Input: P and q two prime numbers.

  1. Choose two very large random prime integers: p and q
  2. Compute n and φ(n):n = pq and φ(n) = (p-1)(q-1)
  3. Choose an integer e, 1 < e <φ(n) such that:

     gcd(e, φ(n)) = 1(where gcd means greatest common denominator)
  4. Compute d, 1 < d <φ(n) such that:ed ≡ 1 (mod φ(n))

     The public key is (n, e) and the private key is (n, d)

     The values of p, q and φ(n) are private.

2. User File Upload Stage

a. *File Level Deduplication:*

After registration ,user becomes ready to upload file on cloud,to upload file there are three algorithmsrunning first for File encryption AES,to generate hash key of file MD5 is used,tosplit file into no of shares we are using Secret sharing schemes who enables more highly reliable and secure level.To upload file the user first make interaction with s-csp to perform deduplication ,then user firsltly compute hash value and sends for file duplicate check.

- If duplicate is found:User computes and sends hash value H(F) of file to the server if the hash value matches then it gives reference pointer to the file which is already present on server, that means user cannot upload same file on cloud.
- If No duplicates Found: He/She runs the secret sharing algorithm SS over F to get {si}=Share(F), where si is i'th shard of F.

ii. *Ramp secret sharing schemes Algorithm*: In this two algorithms present share and recover. The secret is divided and shared by using Share algorithm.RSSS use to secretly splitting of secret into shards. Specifically (n,k,r)-RSSS (where n>k.r>=0)generates n shares from secret.Finally user uploads set of values {Uid, Fid, Fname, Encrypted file, si, H(F)}

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 6, June 2016

b. *Block Level Deduplication*:

To upload file F user needs to firstly perform file level deduplication before uploading his file. If no duplicate is found user divide this particular file into blocks and perform block level deduplication. He/she divides F into set of fragments {Bi}(i=1,2,3……).For each fragment Bi user will perform a block level duplicate check by computing H(Bi).

- If duplicate is found: User computes hash of block by using MD5 and sends to s-csp,with its encrypted block and identifier. If it matches s-csp returns a block pointer of Bi and does not need to upload Bi.
- If No duplicates found: user runs secret sharing algorithm SS over {Bi} and gets {cij}=Share(Bi) where cij is j'th secret share of Bi.Finally user uploads set of values{Uid,Fid,Bi,H(Bi),Encrypted Block,cij} to the server.

3. TPA File Auditing

Once the TPA detects a data corruption during the auditing process, he/she will report the error to users. User sends hash key with tag of file to TPA for integration check Then TPA generates the challenge message CM = {fi,fid,uid,H(F)} by running the Challenge algorithm.

iii. Challenge Algorithm:

chal = {f(i,H(F)} On receiving the challenging message CM, CSP givesresponse of respective information in CM' to TPA.And Finally, TPA runs verification process by matching both CM and CM'.

By verifying both CM and CM', TPA sends response to user regarding file safety.

a) *Algorithm: Proof chal ={i,fid,uid}*

i. Server chooses a _le id (cfid) and user id (cuid) from database wherecfid = fid and cuid = uid
ii. To bind from server computes comparison and get hash value H = H(F)
iii. Send H to TPA.

CSP Proof: With the response from CSP, the TPA runs VerifyProof to validate it by firstcomputing and then checking the verification equation.

b) Algorithm: Verify Proof= H

i. Get H from CSP.
ii. Compare H(F) and
iii. If (H(F)equal to) then
iv. Send response r
v. Else send response r'
vi. Send verification response to user
vii. File Download

4. User File Download stage

When user wants to download file, user request forwarded towards KDC.If user wants to download private filethis request pass to data/file owner in purpose of file safety. After response from file owner, it will allow .for download access .To download file F user needs to download the secret shares {si} of file from k out of n storage server. After getting enough shares ,File F reconstruct by using algorithm RSSS-recover. RecoverAlgorithm :takes input as any k out of n shares and then output original secret S.At last KDC perform decryption by applying AES decryption algorithm by using private key of that user.

IV. RESULT ANALYSIS

Figure 2 shows actual output which get from our proposed system. X axis illustrates number of files and Y axis represents the total database size. It shows actual storage space in database, withfile level duplication having large storage space as compare to block level duplication and it shows comparatively that without deduplication large storage space required than with deduplication schemes. Thus by using deduplication scheme we can reduce storage space overheads.

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 6, June 2016

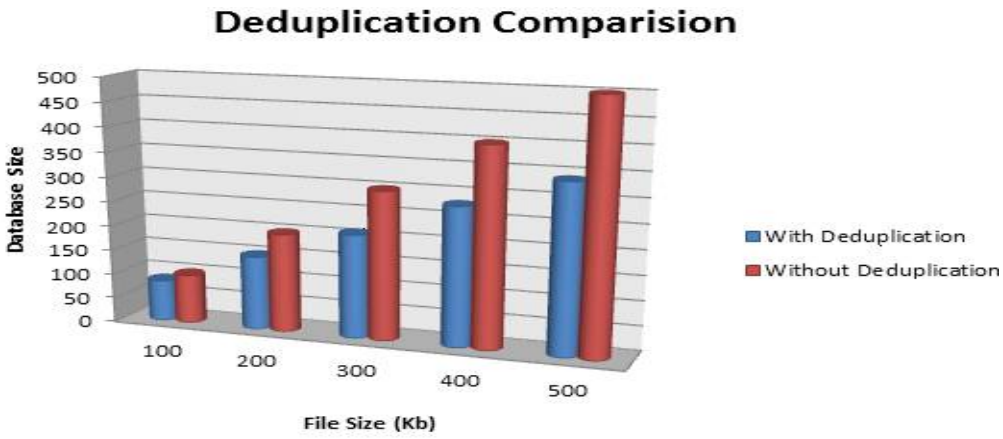

Figure 2.comparison of with deduplication and without deduplication

V. CONCLUSION

We proposed the distributed deduplication system to improve reliability of data by achieving confidentiality of data, without use of encryption mechanism. We have implemented ramp secret sharing scheme.Additionally,in this paper, a secure cloud storage system is proposed which introduce privacy-preserving Third Party Audit, also further extended to enable the TPA to perform audits for multiple users efficiently and in a batchManner. Also this paper introduce (KDC) which store all users key that improves security for data leakage by providing key to authorized user only.

REFERENCES

[1] Jin Li, X. Chen, X. Huang, S. Tang, Y. Xiang, M. M. Hassan and A. Alelaiwi, "Secure Distributed Deduplication Systems with Improved Reliability", IEEE Transactions on Computers, 2015
[2]. 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*, 2014, pp. vol.25(6), pp. 1615–1625.
[3] M. Li, C. Qin, P. P. C. Lee, and J. Li, Convergent dispersal: Toward storage-efficient security in a cloud-of-clouds, inThe 6th USENIX Workshop on Hot Topics in Storage and File Systems, 2014.
[4]"Message-locked encryption and secure deduplication," in *EUROCRYPT*, 2013, pp. 296–312..
[5] D. Harnik, B. Pinkas, and A. Shulman-Peleg, "Side channels in cloud services: Deduplication in cloud storage." *IEEE Security &Privacy*, vol. 8, no. 6, pp. 40–47, 2010.
[6] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, and D. Song, Provable Data Possession atUntrusted Stores, Proc. 14th ACM Conf. Computer and Comm. Security (CCS 07), pp. 598-609, 2007.
[7] M.A. Shah, R. Swaminathan, and M. Baker, Privacy- Preserving Audit and Extraction of Digital Contents, Cryptology ePrint Archive, Report 2008/186, 2008.
[8] M. Bellare, S. Keelveedhi, and T. Ristenpart, Dupless: Serveraided encryption for deduplicated storage, in USENIX Security Symposium, 2013.
[9]A. Shamir, How to share a secret, Commun. ACM, vol. 22, no. 11,pp. 612613, 1979.
[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, 2014, pp. vol. 25(6), pp. 16151625.
[11]Yuan, Jiawei, and Shucheng Yu. "Secure and constant cost public cloud storage auditing with deduplication." Communications and Network Security (CNS), 2013 IEEE Conference on.IEEE, 2013.
[12]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 491500. ACM, 2011.