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

Authorized and Secure De-duplication Based On Hybrid Cloud

Minal Pore, Dr. Roshani Raut (Ade)

M.E. Student, Dept. of Computer Engineering, Dr. D.Y. Patil School of Engineering and Technology, Pune, India

Professor, Dept. of Computer Engineering, Dr. D.Y. Patil School of Engineering and Technology, Pune, India

ABSTRACT: Tremendous increase in number of cloud users and their data volume become critical challenge for cloud computing. To manage such large volume of data cloud providers are using data de-duplication. It is data compression technique which eliminates duplicate copies of repeating data and saves network bandwidth. Along with de-duplication data confidentiality and integrity achievement is also important. De-duplication can be performed on file level as well as block level. To provide more security paper uses both file level and block level data de-duplication. To protect the confidentiality of sensitive data while supporting de-duplication, the convergent encryption technique has been proposed to encrypt the data before outsourcing. To enhance data security and achieve confidentiality this paper formally addresses the problem of authorized data de-duplication. Duplicate check is based on differential privileges of users. Duplicate check is carried on data itself. To retain the privacy of sensitive data concept of hybrid cloud is proposed. Several new de-duplication constructions supporting authorized duplicate check in hybrid cloud architecture are proposed. Security analysis demonstrates that our scheme is secure in terms of the definitions specified in the proposed security model. We show that our proposed authorized duplicate check scheme incurs minimal overhead compared to normal operations.

KEYWORDS: Authorized de-duplication, Cloud computing, Convergent encryption, De-duplication, Hybrid cloud

I. INTRODUCTION

Cloud computing is an evolved computing terminology based on utility and consumption of computing resources. It allows on demand access to computer services, computing resources, remote servers, networks and storage on the basis of pay as you go model. To make the data management compatible and easy data de-duplication is a very important technique used by cloud storage providers. Simple idea behind de-duplication is to store duplicate data only once. It is specialized data compression technique which eliminates the redundant data and improves the storage utilization Redundant data is eliminated by de-duplication and single physical copy of data is kept and when redundant data is referred, it is directed to that copy. Data de- duplication occurs at file level as well as block level. The duplicate copies of identical file eliminate by file level de-duplication. Block level duplication eliminates duplicates blocks of data that occur in non-identical files.

During data de-duplication several security and privacy issues arises, as user's sensitive data are susceptible to inside and outside attackers [12]. Traditional encryption is contradictory with data de-duplication. Traditional encryption requires different users to encrypt their data with own keys which leads to different cipher texts which is not compatible for de-duplication. Hence, data de-duplication works with convergent encryption technique which enforces data confidentiality while making de-duplication feasible. In convergent encryption first, hash value is obtained from the content and hash of the data is considered as a key to encrypt data. Hence, CE will encrypt the identical data into the same cipher text, which enables de-duplication on the cipher text.

To control unauthorized access secure proof of ownership protocol is needed [6]. System runs this protocol when duplicate is found and provides a proof to user who owns the same file. After this instead of uploading same file again path pointer is provided to corresponding user [9][10]. In this paper, aiming at efficiently solving the problem of de-duplication with differential privileges in cloud computing, we consider a hybrid cloud architecture consisting of a public cloud and a private cloud. Here, private cloud is considered as proxy to allow data owner/user to securely perform duplicate check. Such architecture is practical and has attracted much attention from researchers. Data owners can outsource their data only by utilizing public cloud and data operation is managed by private cloud. A new de-duplication system supporting differential duplicate check is proposed under this hybrid cloud architecture where the

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

storage cloud sever resides in the public cloud. The user is only allowed to perform the duplicate check for files marked with the corresponding privileges. Furthermore, we enhance our system in security. Specifically, we present an advanced scheme to support stronger security by encrypting the file with differential privilege keys. In this way, the users without corresponding privileges cannot perform the duplicate check. Furthermore, such unauthorized users cannot decrypt the cipher text even collude with the S-CSP.

II. RELATED WORK

JinLiandYanKitLipresentedhybridcloudapproachforsecureauthorizedde- Jin Li and Yan Kit Li presented hybrid cloud approach for secure authorized de-duplication. It aims for solving the problem of the de-duplication with different privileges in cloud computing. [1] Jin Li proposed secure de-duplication with efficient and reliable convergent key management contain the different techniques which is used in the secure de-duplication and remove the duplicate copies of data for reduce the storage space in cloud system. For that purpose, use the convergent encryption to providing the data confidentiality and encrypt/decrypt a data copy with a convergent key, which is given by computing the cryptographic hash value of the content of data copy itself. This technique is used for reduce the storage space and bandwidth also provide the confidentiality. [2]

M. Bellare, S. Keelveedhi, and T. Ristenpar proposes DupLess: Server aided encryption for de-duplicated storage for cloud storage service provider like Mozy, Dropbox, and others perform de-duplication to save space by only storing one copy of each file uploaded. Message lock encryption is used to resolve the problem of clients encrypt their file however the saving is lock. Dupless is used to provide secure de-duplicated storage as well as storage resisting brute-force attacks [3].

Twin clouds: Architecture for secure cloud computing proposed Client uses the trusted Cloud as a proxy that provides a clearly defined interface to manage the outsourced data, programs, and queries. It stores large amount of data and low latency [7].

S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg pre- sented Proof of Ownership for check the ownership of user who having the authority for access the file or uploaded data from the given cloud storage system. Sometimes user not upload the file but it tries to access the data from the given cloud to avoid this problem use proof of ownership algorithm Used to provide authorized access to user[8].

For the cloud storage de-duplication system Yuan et al. [11] proposed an integrity check method which decreases the storage size of the tags.

III. RELEVANT MATHEMATICS

First, paper presents attempt with token generation technique TagGen (F, kp). Basic idea is to issue the privilege keys to corresponding users who will compute tags for duplicate check which is based on privilege keys and files.
Assume that there are N users in the system and the privileges in the universe are defined as P = {p1. . . ps}. For each privilege p in P, a private key kp will be selected. User having set of privileges PU will be assigned the set of key, $\{k_{pi}\}_{pi \in Pu.}$
If data owner U with privilege Pu wants to upload the file on cloud storage server and share that file with the users having privilege set Pf={Pj}.
Here, first user generates the file token ϕ f for file f such that ϕ f, p = Tag Gen (F, kp) for all
$p \in P_f$

While performing duplicate check if server cloud storage provider (S-CSP) found the duplicate, the user has to proceeds proof of ownership for that particular file with the S-CSP. After successful verification the user will be assigned a pointer, which allows him to access the file. Otherwise, if no duplicate is found, the user encrypts the file CF = EncCE(kF, F) with the convergent keykF = KeyGenCE(F) and uploads (CF,{ϕ'f,p}) to the cloud server. The convergent key kf is stored by the user locally.

- Set Theory:

Let S represents proposed system such that
S={s, e, X,F_s, Y,D,N,F_r S_f,TPA,|Φ}
Where,
s = F_{upload}:F□C
e= F_{store} :U□C(L

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

D= Already present file.
N= New file.
F_r =Fragmentation.
TPA= Third Party Auditor.
S_f = Fragments encoded using Secret Sharing Scheme.
Input:
X = {F}.
Where,
F = F_1, F_2,….. F_n, n number of files.

IV. PROPOSED SYSTEM

A. *SYSTEM ARCHITECTURE:*

System defines three main entities namely data owner, private cloud and S-CSP which resides in public cloud. User/Data owner uploads data to cloud storage and accesses it later. User can upload only unique data. Each user is assigned with privileges for authorized de-duplication. As public cloud is not fully trusted, private cloud provides data owners with an execution environment and infrastructure working as an interface between user and the public cloud. Tokens required for duplicate check are managed by private cloud in our system. S-CSP is a storage server resides in public cloud and responsible for outsourcing services and data on behalf of data owner. De-duplication process is carried out in S-CSP.

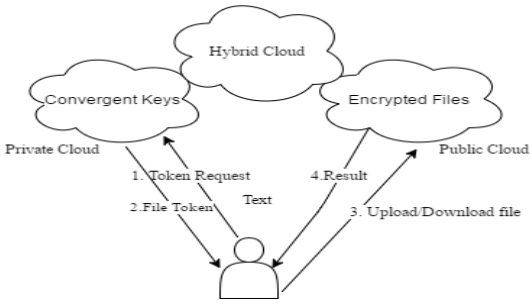

Fig.1. System Architecture (Authorized De-duplication)

B. *DESCRIPTION OF THE PROPOSED ALGORITHM:*

1. Convergent Encryption(SHA-256):

Along with de-duplication to support data confidentiality system is using convergent encryption. A convergent encryption scheme can be defined with four primitive functions:

- KeyGenCE (M)! K is the key generation algorithm that maps a data copy M to a convergent key K;
- EncCE (K, M)! C is the symmetric encryption algorithm that takes both the convergent key K and the data copy M as inputs and then outputs a cipher text C;
- DecCE (K, C)! M is the decryption algorithm that takes both the cipher text C and the convergent key K as inputs and then outputs the original data copy M; and
- TagGen (M)! T (M) is the tag generation algorithm that maps the original data copy M and outputs a tag T (M).

2. Proof Of Ownership

To perform duplicate check S-CSP runs PoW algorithm. It enables users to prove their ownership of data copies to the storage server. PoW is interactive algorithm run by prover. The verifier derives a short value ϕ (M) from a data copy M. To prove the ownership of the data copy M, the prover needs to send ϕ to the verifier such that $\phi' = \phi$ (M).

PSEUDO CODE

Step1: Calculate the two convergent key values
Step2: Compare the two keys and files get accessed.
Step3: Apply de-duplication to eradicate the duplicate values.
Step4: If any other than the duplicates it will be checked once again and make the data unique.

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

Step5: That data will be unique and also more confidential the authorized can access and data is stored.

3. Shamir's Secret Sharing Scheme:-

Secret sharing refers to the method of distributing secret amongst the group of participants, each of which is allocated share of secret. The secret can only be reconstructed when the shares are combined together.

- Shamir's Secret Sharing:

Based on polynomial interpolation

-k points are needed to fully define a unique polynomial of degree k-1 –

-Example: 2 points fully define straight line

3 points fully define quadratic and so on...

It's (k, n) threshold scheme

-Dealer D distribute a secret s to n players

-At least k participants are required to construct a secret s

4. Message Authentication(HMAC-SHA1)

This Algorithm is most useful in cloud computing as it is the arising technology to minimize the user burden in the updating of data in business using internet. Instead of local data storage and maintenance, the user is assisted with the cloud storage so that the user can remotely store their data and enjoy the on-demand high quality application from a shared pool of resources. The data stored must be protected in the cloud storage. Proposed system has mentioned a method that uses the keyed Hash Message Authentication Code (HMAC) security.

5. Fragment Allocation Algorithm(T-coloring Graph Technique):

T coloring consists of assigning a color c(v) , v∈V so that the absolute difference of assigned colors to adjacent nodes do not belong to any number in T, namely if (x,y) ∈E ⇒|c(x)-c(y) ∉T.

Algorithm shows placement scenario of fragments

6. Auditing Algorithm:

The cloud servers and TPA interact with one another to take a random sample on the blocks and check the data intactness in this procedure. This algorithm is performed by the TPA with the information of the file as input and a challenge as output. Taking the proof, public parameter and the corresponding challenge *C* as input, it outputs 1 if the verification passed and 0 otherwise.

V. RESULTS AND DISCUSSION

A few new de-duplication developments supporting approved copy check in half and half cloud engineering, in which the copy check tokens of documents are produced by the private cloud server with private keys. Security examination exhibits that our plans are secure as far as insider and outsider assaults indicated in the proposed security model. As a proof of idea, our proposed system approved copy check plan and direct proving ground investigates our model. System demonstrated that our approved copy check plan acquires insignificant overhead contrasted with united encryption and system exchange.

- FILE UPLOADING:

Figure shows time required to upload a file. This graph computes the time required to upload particular file by data owner in milliseconds. User can upload text or PDF files.

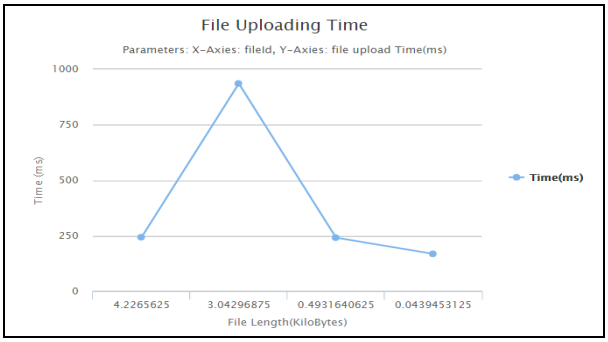

Fig.2. File uploading graph

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

- FILE DOWNLOADING:

Following graph shows time required to download particular file by data owner. File size is considered in kilobytes and time required is in milliseconds.

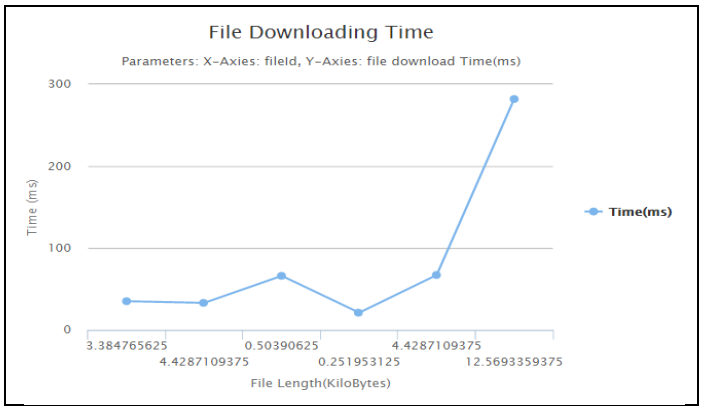

Fig. 3 File downloading graph

- STORAGE CONSUMPTION:

Fig.4 is a graph showing how space utilization is reduced after applying de-duplication technique on storage. Graph describes memory saved after performing de-duplication.

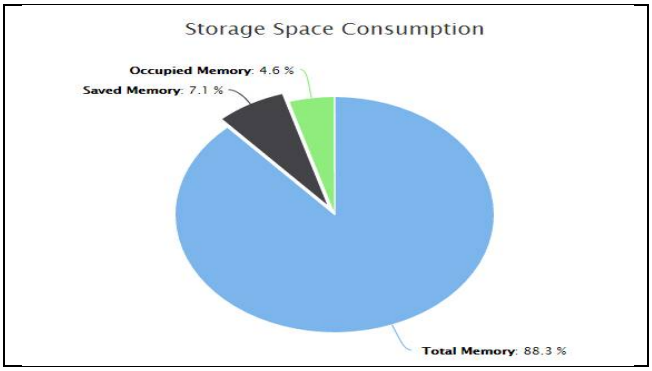

Fig. 4 Storage space consumption graph

III. SYSTEM ANALYSIS AND EVALUATION

Proposed system achieves data confidentiality by encrypting data before outsourcing to the storage server. To achieve more security system stores the file blocks by T-coloring which achieves high security. System analyses the computational complexity of the two most important operations: storage and retrieval. N is the mean number of blocks per file and M the total number of blocks in the system.

Table 1 Complexity (Storage & Retrieval)

	Storage	Retrieval
Encryption	$O(N)$	$O(N)$
Hash	$O(N)$	$O(N)$
Lookup in data structures	$O(N \log M)$	$O(N)$
Other	$O(N)$	$O(N)$

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

IV. CONCLUSION AND FUTURE WORK

Along with secure de-duplication, notion of authorized de-duplication is proposed to protect data security. In duplicate check, differential privileges of users are considered. We address the problem of privacy preserving de-duplication in cloud computing and propose a new de- duplication system supporting for Differential Authorization and Authorized Duplicate Check. To better protect data security, this paper makes the first attempt to formally address the problem of authorized data de-duplication. Different from traditional de- duplication systems, the differential privileges of users are further considered in duplicate check besides the data itself. We also present several new de-duplication constructions supporting authorized duplicate check in hybrid cloud architecture. Security analysis demonstrates that our scheme is secure in terms of the definitions specified in the proposed security model. As a proof of concept, we implement a prototype of our proposed authorized duplicate check scheme and conduct tested experiments using our prototype. We show that our proposed authorized duplicate check scheme incurs minimal overhead compared to normal operations.

REFERENCES

1. Jin Li and Yan Kit Li A˙ Hybrid cloud approach for secure authorized de-duplication, IEEE Transaction on parallel and distributed system,vol:pp:99 2014.
2. J. Li, X. Chen, M. Li, J. Li, P. Lee, and W. Lou. Secure de-duplication with efficient and reliable convergent key management. In IEEE Transactions on Parallel and Distributed Systems, 2013.
3. M. Bellare, S. Keelveedhi, and T. Ristenpart. Dupless: Server aided encryption for deduplicated storage. In USENIX Security Symposium, 2013.
4. M. Bellare, S. Keelveedhi, and T. Ristenpart. Message-locked encryption and secure deduplication. In EUROCRYPT, pages 296 312, 2013.
5. J. Xu, E.-C. Chang, and J. Zhou, Weak leakage-resilient client- side deduplication of encrypted data in cloud storage, In ASIACCS, pages 195-206, 2013.
6. W. K. Ng, Y. Wen, and H. Zhu. Private data deduplication protocols in cloud storage. In S Ossowski and P. 2012.
7. S. Bugiel, S. Nurnberger, A. Sadeghi, and T. Schneider. Twin clouds: An architecture for secure cloud computing. In Workshop on Cryptography and Security in Clouds (WCSC 2011), 2011.
8. 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.
9. J. Yuan and S. Yu., Secure and constant cost public cloud storage auditing with de-duplication, IACR Cryptology ePrint Archive, 2013:149, 2013.
10. J. Stanek, A. Sorniotti, E. Androulaki, and L. Kencl, A secure data de-duplication scheme for cloud storage, In Technical Report 2013.
11. J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M.Theimer, Reclaiming space from duplicate files in a serverless distributed file system, In ICDCS, Harlow, pages 617-624, 2002.
12. P. Anderson and L. Zhang, Fast and secure laptop backups with encrypted de-duplication, In Proc. of USENIX LISA, 2010.
13. A. Rahumed, H. C. H. Chen, Y. Tang, P. P. C. Lee, and J. C. S. Lui, A Secure cloud backup system with assured deletion and version control, In 3rd International Workshop on Security in Cloud Computing, 2011.
14. M. W. Storer, K. Greenan, D. D. E. Long, and E. L. Miller, Secure data de-duplication, In Proc. of StorageSS, 2008.
15. Z. Wilcox-OHearn and B. Warner., Tahoe: the least- authority file system, In Proc. of ACM StorageSS. 2008.
16. Zhang, X. Zhou, Y. Chen, X. Wang, and Y. Ruan, Sedic: privacy aware data intensive computing on hybrid clouds, In Proceedings of the 18th ACM conference on Computer and communications security, CCS11, pages 515-526, New York, NY, USA, 2011.

BIOGRAPHY

Miss.Minal Pore received B.E degree in computer science and engineering from Shivaji University, Kolhapur in 2012. She is currently pursuing Master's degree in computer networks from Dr.D.Y Patil school of engineering and technology, Pune under Savitribai Phule Pune University. Her research interest includes Cloud computing and networking.