IJCSE International Journal of Computer Sciences and Engineering Open Access

Research Paper Volume-4, Special Issue-4, June 2016 E-ISSN: 2347-2693

De-duplication with Authorization in Hybrid Cloud Approach for Security

Guljar P. Shaikh

Department of Information Technology, India

Available online at: www.ijcseonline.org

Received: May/26/2016 Revised: Jun/02/2016 Accepted: Jun/12/2016 Published: Jun/30/2016

Abstract— In cloud computing, most of the communication is done in consideration of file processing, and therefore it turn out to be very critical and important to deliver competent method for data security. The main focus in on data deduplication to provide security services over cloud for stored files. Data deduplication is data compression that is used to eliminate the matching copies of echoing data. This procedure is frequently used for reducing the storage space and save bandwidth over cloud. The term deduplication used for data security and privacy offers to the stored data files, Also encryption methods are applied to stored file information for more security purpose. In this paper we have studied the authorized data deduplication technique for providing security by using user's differential privileges. Novel deduplication structures are offered for supporting authorized duplicate check. This paper shows how the security is achieved in hybrid cloud throughout the data deduplication process.

Keywords— *Data Security,Data Deduplication,, Privilege authorization, Hybrid Cloud.*

I. INTRODUCTION

The cloud helps to get various information from any part of the world at any time. Data management is measurable in cloud computing, deduplication has been a significant technique recently used. Data deduplication is expert data compression method for eliminating replica copies in cloud[1]. The method helps to improve storage utilization and can also be used in network data transfers to decrease the number of bytes that can be sent. Instead of having numerous data having similar content, deduplication disregards repeated data by keeping only one physical copy and referring other repeated data to that copy. Deduplication is mainly two types namely the file level and the block level[4]. For file-level deduplication, it makes the elimination of replica copies of the same file. For the block level, it makes elimination of replica blocks of data that happens in non-identical files[2]. As the encryption operation is done in deterministic style and is got from the data content, matching data copies will be generating the same convergent key[4] and also the same cipher text. The user will find out a duplicate of the file only if there is a copy of the file and a matched privilege which is stored in cloud. To save cost and efficiently manage, the data will be moved to the storage server provider in the public cloud with particular privileges. Deduplication method will apply on stored only one copy of the similar file.

II. EXISTING SYSTEM

Data deduplication systems, the private cloud is treated as a proxy to authority knowledge owner to decisively achieve duplicate deliberate with differential privileges. These kinds of designs are sensible and have concerned abundant consideration from researchers. The previous de-duplication system does not support the differential authorization and the duplicate check, which is an important security aspect in many of the applications. In the authorized de-duplication system each user is assigned with set of privileges during the system start up [14]. Each file uploaded to the cloud is associated with a set of privileges assigned, which specifies what kind of user is allowed to use the duplicate check and access the files. Before accepting a duplicate check request for a file, the user needs to take his/her file and his/her own privileges as inputs [6]. Then the user can actually find the duplicate for this file, if only there is a copy of these files that is matched with the privilege stored in the cloud.

Limitation

- Confidentiality is incompatible with deduplication.
- Identical information copies totally different of various users can result in different cipher texts, creating deduplication not possible

III. PROPOSED SYSTEM

In this paper, we have a propensity to enhance and improve our system in security aspects. The novel deduplication structure supports stronger security to system by encrypting the file with giving differential rights for accessing stored files and secrete keys. Unauthorized Users are not having rights; cannot execute the duplicate check on file. Proof of ownership is supported for providing more confidentiality to the stored data. Analysis of the security how that the system is secure in terms of the explanations put down in the designed security model[1].

Advantages

- Only authorized users are permitted achieve the duplicate check
- Reduce the storage space
- Improve network bandwidth
- Support stronger security by encrypting the file.
- Reductions in costing

IV. SYSTEM ARCHITECTURE

The following figure 1.shows the system architecture

1. Users - A user is an object who wants to outsource data to storage over Cloud named SCSP and access that stored data later on whenever needed. In storage system, in support of data deduplication, the user can able to upload only unique data files and not any matching data to the cloud. This leads in saving network bandwidth also the storage area[1][12].

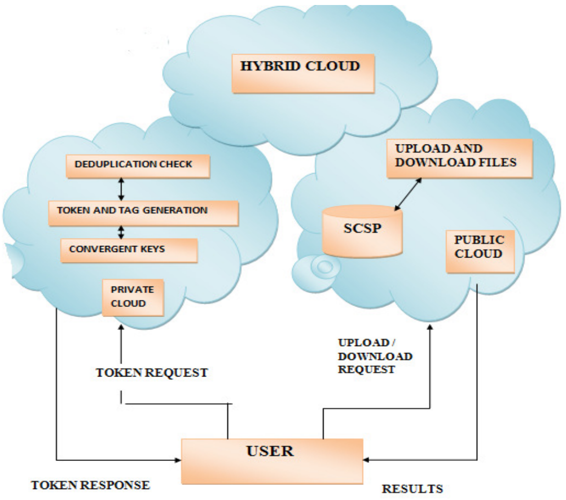


Figure 1 Architecture of proposed Work[1]

2. Private Cloud - Private cloud is able to offers data user with an implementation situation and organization working. This is act as interface amongst the user and the public cloud. The secret keys operations are managed by private cloud. This also allows user to upload files and submit their queries to be securely stored and computed respectively[1][2].

3. S-CSP: This entity is one of the recognized parts into the public cloud. It is basically used for offering services to outsource the data. S-CSP achieves deduplication for storing only the distinct data rather than storing the duplicate data[1][2].

V. PROPOSED ALGORITHM

```
Input: File F
Output: Encrypted file stored over cloud.
Steps:
1.  Begin
2.  The object to be encrypted is validated to ensure it is suitable for this type of encryption.
3.  hash value of data is created (e.g.HMAC-SHA1)
4.  hash (key)←Enc(File{Contents}) (e.g. AES-CBC)
5.  SCSP← Hash (Key)
6.  If (File1{Contents}==File2{Contents})
7.          Request Denied
8.  Else
9.          User←File{Key}
10.         SCSP←User(File{Key})
11.         Enc(File)
12.         Upload File to Cloud
13. End if
14. End
```

VI. MATHEMATICAL MODEL

Let C be the set of clients that is it may have one or more clients. CSP is the cloud service provider. GUI is the graphical user interface. F denotes a file. Hence, the system S is given as follows:

S = {C, CSP, GUI}

Global set of entities involved are described below:

G ={D, U, PC, PRC}

Where,

D= data owner,
U=End user,
PC=Public cloud,
PRC=Private cloud.

There are two functions p1 and p2 which are described as verification of clients and verification of end users respectively.

The processes that are carried out are: Verification, Encryption, Decryption and Storage. In the verification process, the user needs to login with a username and password. The inputs to the encryption process are the contents to encrypt and the key. Similarly, the inputs to the decryption process are key and the cipher text to be decrypted. Storage contains H(C).

Process:

Verification[Y/N] = Login (Username, Password)

Encryption = (Key, Content)

Decryption = (key, Cipher)

Storage = H (C)

VII. RESULT AND ANALYSIS

We have performed the given experiment on dot net platform with SQL as a database to keep track of users and files. The system is evaluated with the windows machine with latest configuration with 4GB RAM and I series processor. We have uploaded files with various formats and noted down different execution time.

The following parameters are checked

A. Execution Time:

The overall execution time for file upload including de-duplication check and encryption for particular file is as mentioned in below table1. The graph analysis for execution time shown in figure2.

File Size	Upload File (seconds)	Download File (Seconds)
10bytes	16	0
1 kb	19	3
10kb	22	1
100kb	25	6
1mb	27	7

Table 1.Execution time

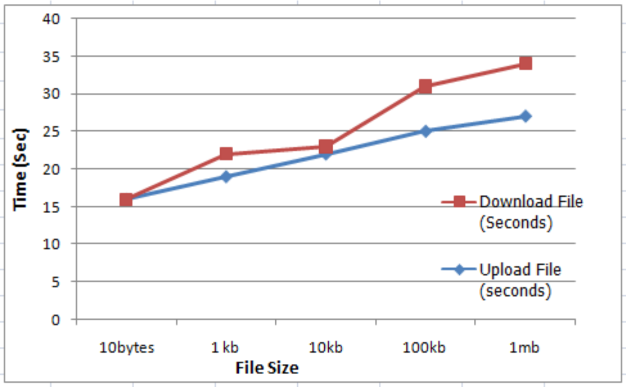


Figure 2.Anlaysis of Execution time

B. Throughput:

Based on the execution time and file size we have also calculated the throughput of the system and which clearly shows that our system improves throughput with cryptographic tuning and domain separation. Throughput of proposed system is as shown in figure 3.

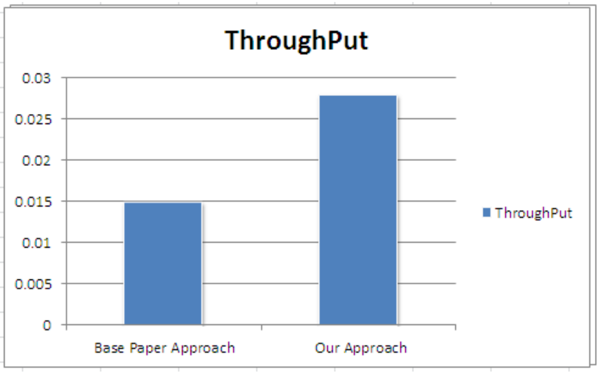


Figure 3.Anlaysis of Throughput

VIII. CONCLUSION

The notion of authorized data deduplication was proposed to protect the data security by providing duplicate check. We also presented hybrid cloud architecture, in which the duplicate-check tokens of files are generated by the private cloud server with secret keys and response given to users. We have used convergent encryption which deals with brute force attack by using Cryptographic to make better authorized deduplication technique.

REFERENCES

[1] Guljar P. Shaikh,S. D. Chaudhary, PriyankaPaygude and Debnath Bhattacharyya, “Achieving Secure Deduplication by Using Private Cloud and Public Cloud”, In International Journal of Security and Its Applications Vol. 10, No. 5, 2016, pp.17-26.

[2] Jin Li,yan kit li,xiaofengchen,patrickP.C.lee and wenjinglou,“A Hybrid Cloud Approach for Secure Authorized Deduplication”, In IEEE Transactions on Parallel and Distributed Systems, April 7,2015,pp 1206 – 1216, DOI:10.1109/TPDS.2014.2318320.

[3] MihirBellare, SriramKeelveedhi, and Thomas Ristenpart, “Dupless: Serveraided encryption for deduplicated storage”, In USENIX Security Symposium, Washington DC, August 14-16, 2013, pp 179-194.

[4] MihirBellare, SriramKeelveedhi, and Thomas Ristenpart, “Message-locked encryption and secure deduplication”, in proceedings of EUROCRYPT, Athens Greece, March 3, 2013, pp 296–312.

[5] ShaiHalevi, Danny Harnik, Benny Pinkas, and Alexandra Shulman-Peleg, “Proofs of ownership in remote storage systems”. In Proceedings of the 18thACM Conference on Computer and Communications Security, Hangzhou, China, 2011,pp 491–500.

[6] S. Bugiel, S. Nurnberger, A. Sadeghi, and T. Schneider. “Twin clouds: An architecture for secure cloud computing”. In Workshopon Cryptography and Security in Clouds (WCSC 2011), 2011

[7] Jin Li, XiaofengChen,MingqiangLi,Jingwei Li and Patrick P. C. Lee, "Secure Deduplication with Efficient and Reliable Convergent Key Management", In IEEE Transactions on Parallel and Distributed Systems, May 12, 2014.pp 1615 – 1625, DOI: 10.1109/TPDS.2013.284.

[8] J. Yuan and S. Yu. "Secure and constant cost public cloud storage auditing with deduplication". IACR Cryptology ePrint Archive, 2013:149, 2013.

[9] M. Bellare, C. Namprempre, and G. Neven, "Security proofs foridentity-based identification and signature schemes," J. Cryptol.,vol. 22, no. 1, pp. 1–61, 2009.

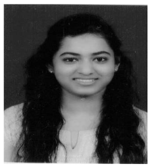
[10] P. Anderson and L. Zhang, "Fast and secure laptop backups with encrypted de-duplication," in Proc. 24th International Conference Large Installation System Admin., 2010, pp. 29–40.

[11] W. K. Ng, Y. Wen, and H. Zhu, "Private data deduplication protocols in cloud storage," in Proc. 27th Annual ACM Symp. Appl. Computer, 2012, pp. 441–446.

[12] Qinlu He, Zhanhuai Li, Xiao Zhang, "Data Deduplication Techniques", in International Conference on Future Information Technology and Management Engineering, China, 2010, pp.43-433.

[13] D. Meister and A. Brinkmann, "Multi-Level Comparison of Data Deduplication in a Backup Scenario", SYSTOR '09 Proceedings of SYSTOR 2009: The Israeli Experimental Systems Conference, ISBN: 978-1-60558-623-6, Article No. 8, ACM New York, NY, USA ©2009.

Authors Profile



Guljar Shaikh has done her Master of information Technology and currently working as a lecturer at Zeal Polytechnic College. Area of Interests is Cloud Computing, Security and Dot Net Programming.