ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 4, pp: (335-339), Month: October - December 2015, Available at: www.researchpublish.com

Deduplication by Means of Economical and Consistent on Secure Clouds

[1]SUDHARANI CHENNUPALLI, [2]S.UMAMAHESWARA RAO

[1]Student of M. Tech (CSE) [2]Associate.Prof, Department of Computer Science and Engineering,
CHIRALA ENGINEERING COLLEGE, CHIRALA, INDIA

Abstract: **We demonstrate that our proposed approved duplicate check plan causes negligible overhead contrasted with typical operations. To secure the privacy of delicate information while supporting Deduplication, the merged encryption system has been proposed to encode the information before outsourcing. Not quite the same as conventional Deduplication frameworks, the differential benefits of clients are further considered in duplicate check other than the information itself. As a proof of idea, we actualize a model of our proposed approved duplicate check plan and behavior test bed examinations utilizing our model. Security investigation exhibits that our plan is secure as far as the definitions indicated in the proposed security model. Information Deduplication is one of essential information pressure procedures for dispensing with duplicate duplicates of rehashing information, and has been broadly utilized as a part of distributed storage to lessen the measure of storage room and spare data transmission. We likewise show a few new Deduplication developments supporting approved duplicate weigh in a half and half cloud structural planning. To better Ensure information security, this paper makes the first endeavor to formally address the issue of approved information Deduplication.**

Keywords: **Deduplication, authorized duplicate check, confidentiality, hybrid cloud.**

I. INTRODUCTION

Cloud computing is internet-based computing where large groups of remote servers are connected to each other to allow the centralized data storage, and online access to computer services or resources. There are three types of cloud - public cloud, private cloud and hybrid cloud. In public cloud, applications and storage are available over the internet for general use. In private cloud, a virtualized data center is used that operates within a firewall. In this research introduce hybrid cloud which is mix of public and private cloud. Cloud computing focused on maximizing the effectiveness of the shared resources. It provides computation and storage resources on the Internet. Cloud resources are usually shared by multiple users and also it is dynamically reallocated per demand. By using cloud computing, many numbers of users can access a single server to retrieve and update their data without purchasing licenses for different applications. Exponential growth of ever increasing data over cloud is became a critical challenge. To face that challenge data deduplication technique is introduced. Data deduplication is data compression technique which eliminate repeated data. Instead of taking multiple numbers of copies of same data, it saves just one copy of the data and other copies are replaced with pointers that lead back to the original copy. It improves bandwidth efficiency and storage utilization. Data dedupliaction protects confidentiality of data. Data deduplication work with convergent encryption technique to encrypt the data before uploading, and we enhance some hashing algorithm which makes the technique very secure before uploading encrypted file into the cloud.

II. SYSTEM MODEL

Hybrid Architecture for Secure Deduplication:

At a high level, our setting of interest is an enterprise network, consisting of a group of affiliated clients (for example,employees of a company) who will use the S-CSP and store data with deduplication technique. The S-CSP

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 4, pp: (335-339), Month: October - December 2015, Available at: www.researchpublish.com

performs deduplication by checking if the contents of two files are the same and stores only one of them. Each privilege is represented in the form of a short message called *token*.

 S-CSP: This is an entity that provides a data storage service in public cloud. The S-CSP provides the data outsourcingservice and stores data on behalf of the users.

 Data Users: A user is an entity that wants to outsource data storage to the S-CSP and access the data later. In a storage system supporting deduplication, the user only uploads unique data but does not upload any duplicate data to save the upload bandwidth, which may be owned by the same use or different users. Every single file is protected with the convergent encryption key and privilege keys to realize the authorized deduplication with differential privileges.

 Private Cloud: Compared with the traditional deduplication architecture in cloud computing, this is a new entity introduced for facilitating user's secure usage of cloud service. Private Keys are managed by private cloud in order to give them privileges as per their designation.

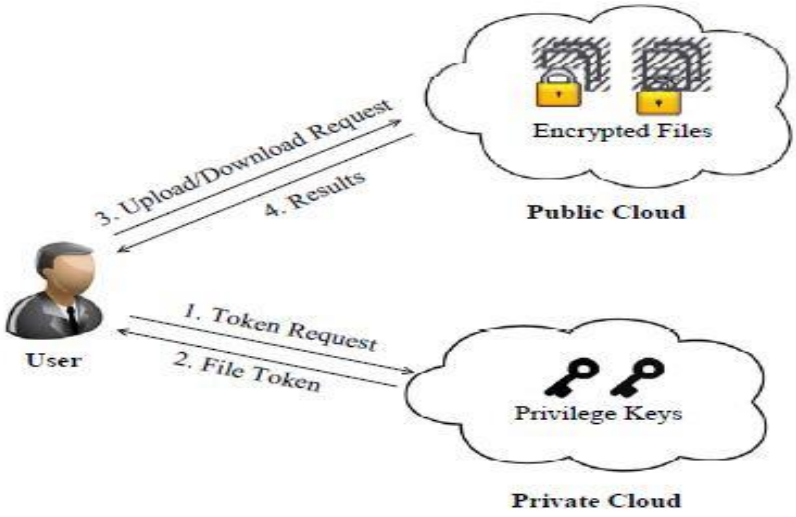


Fig. 1. Architecture for Authorized Deduplication

III. DESIGN GOALS

In this paper, we address the problem of privacy preserving deduplication in cloud computing and propose a new deduplication system supporting for:

 Differential Authorization: Each authorized user is able to access its individual token of his file to perform duplicate check based on authority.Under this assumption, any user cannot generate a token for duplicate check out of his access or without the aid from the private cloud server.

 Authorized Duplicate Check: Authorized user is able to access his/her own token from private cloud,while the public cloud performs duplicate check directly and tells the user if there is any duplicate. The security requirements considered in this paper lie in two folds, including the security of file token and security of data files. For the security of file token, two aspects are defined as unforgeability and indistinguishability of file token. The details are given below.

 Unforgeability of file token/duplicate-check token: User make registration in private cloud for generating file token.Using respective file token he/she upload or download files on public cloud. The users are not allowed to collude with the public cloud server to break the unforgeability of file tokens. In our system, the S-CSP is honest but curious and will honestly perform the duplicate check upon receiving the duplicate request from users. The duplicate check token of users should be issued from the private cloud server in our scheme.

Indistinguishabilityof file token/duplicate-check token:

It requires that any user without querying the private cloud server for some file token, he cannot get any useful information from the token, which includes the file information and key information.

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 4, pp: (335-339), Month: October - December 2015, Available at: www.researchpublish.com

 Data Confidentiality: Unauthorized users without appropriate token , including the S-CSP and the private cloud server, should be prevented from access to the underlying plaintext stored at S-CSP. In another word, the goal of the adversary is to retrieve and recover the files that do not belong to them. In our system, compared to the previous definition of data confidentiality based on convergent encryption, a higher level confidentiality is defined and achieved.

IV. SECURITY ANALYSIS

Proposed system has been designed to solve the differential privilege problem in secure deduplication. The security will be analyzed in terms of two aspects, that is, the authorization of duplicate check and the confidentiality of data.

Security of Duplicate-Check Token:

We consider several types of privacy we need protect, that is, i) unforgeability of duplicate-check token: There are two types' ofadversaries, that is, external adversary and internal adversary. As shown below, the external adversary can be viewed as an internal adversary without any privilege. If a user has privilege p, it requires that the adversary cannot forge and output a valid duplicate token with any other privilege p' on any file F, where p does not match p'. Furthermore, it also requires that if the adversary does not make a request of token with its own privilege from private cloud server, it cannot forget and output a valid duplicate token with p on any F that has been queried. The internal adversaries have more attack power than the external adversaries and thus we only need to consider the security against the internal attacker, ii) indistinguishability of duplicate check token : this property is also defined in terms of two aspects as the definition of unforgeability. First, if a user has privilege p, given a token ϕ', it requires that the adversary cannot distinguish which privilege or file in the token if p does not match p'.

V. PROPOSED SYSTEM

In our system we implement a project that includes the public cloud and the private cloud and also the hybrid cloud which is a combination of the both public cloud and private cloud. In general by if we used the public cloud we can't provide the security to our private data and hence our private data will be loss.So that we have to provide the security to our data for that we make a use of private cloud also. When we use a private clouds the greater security can be provided.In this system we also provides the data deduplication . which is used to avoid the duplicate copies of data.User can upload and download the files from public cloud but private cloudprovides the security for that data.that means only the authorized person can upload and download the files from the public cloud.for that user generates the key and storeed that key onto the private cloud at the time of downloading user request to the private cloud for key and then access that Particular file.

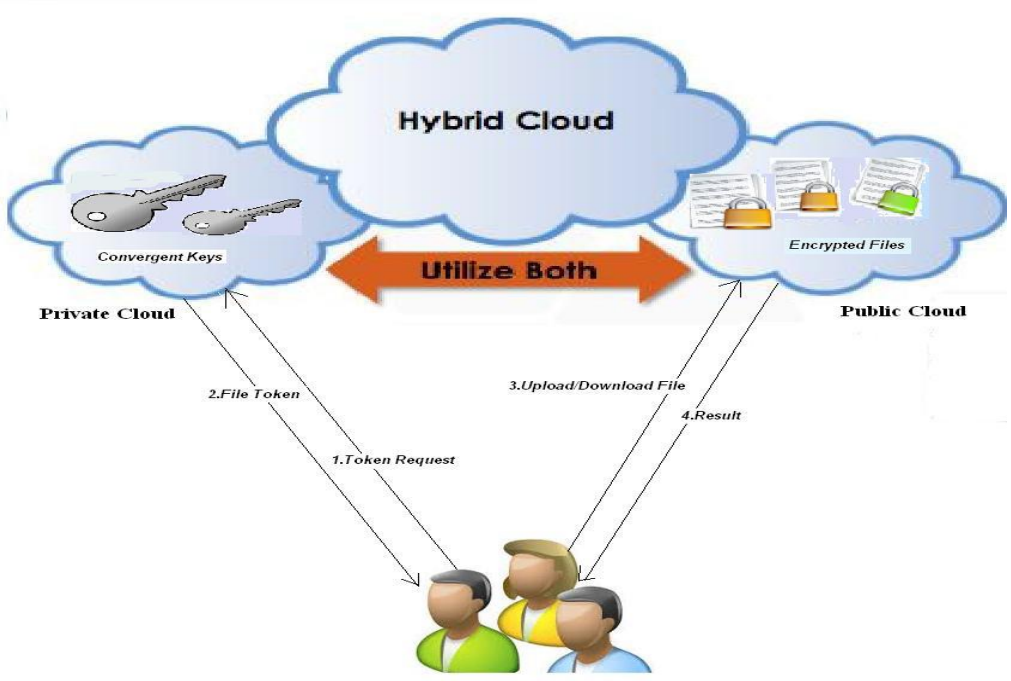


Fig2: Architecture of Authorized Deduplication

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 3, Issue 4, pp: (335-339), Month: October - December 2015, Available at: www.researchpublish.com

VI. ASSESSMENT

Our evaluation focuses on comparing the overhead induced by authorization steps, including file token generation and share token generation, against the convergent encryption and file upload steps. We evaluate the overhead by varying different factors, including 1) File Size 2) Number of Stored Files 3) Deduplication Ratio 4) Privilege Set Size. We break down the upload process into 6 steps, 1) Tagging 2) Token Generation 3) Duplicate Check 4) Share Token Generation 5) Encryption 6) Transfer . For each step, we record the start and end time of it and therefore obtain the breakdown of the total time spent. We present the average time taken in each data set in the figures

File Size:

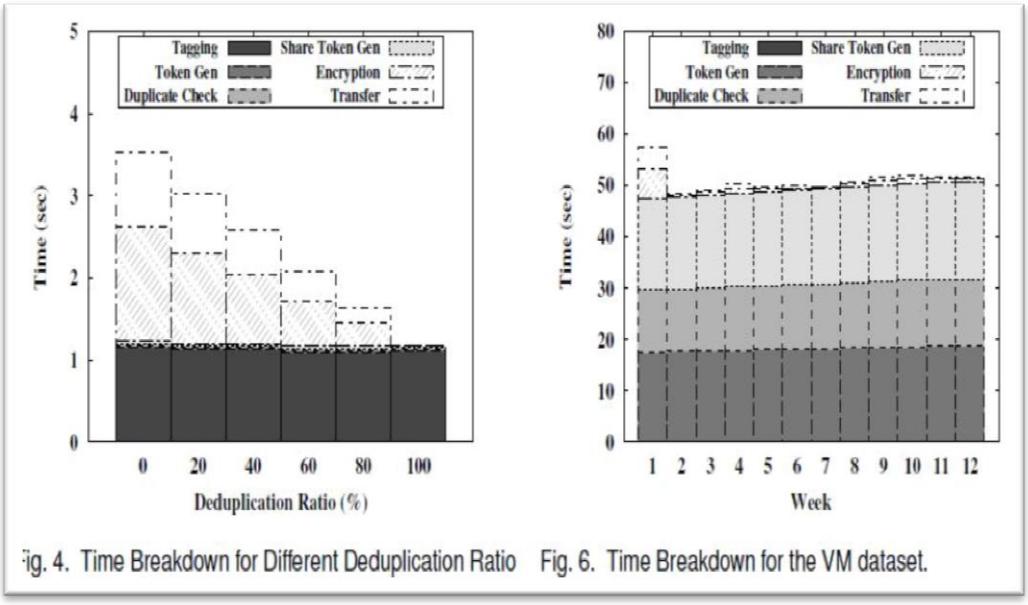
To evaluate the effect of file size to the time spent on different steps, we upload 100 unique files (i.e., without any deduplication opportunity) of particular file size and record the time break down. Using the unique files enables us to evaluate the worst-case scenario where we have to upload all file data. The average time of the steps from test sets of different file size are plotted in Figure 2. The time spent on tagging, encryption, upload increases linearly with the file size, since these operations involve the actual file data and incur file I/O with the whole file.

Number of Stored Files:

To evaluate the effect of number of stored files in the system, we upload 10000 10MB unique files to the system and record the breakdown for every file upload. From Figure 3, every step remains constant along the time. Token checking is done with ahash table and a linear search would be carried out in case of collision.

Deduplication Ratio:

To evaluate the effect of the deduplication ratio, we prepare two unique data sets, each of which consists of 50 100MB files. We first upload the first set as an initial upload. For the second upload, we pick a portion of 50 files, according to the given deduplication ratio, from the initial set as duplicate files and remaining files from the second set as unique files. The average time of uploading the second set is presented in Figure.



ig. 4. Time Breakdown for Different Deduplication Ratio Fig. 6. Time Breakdown for the VM dataset.

VII. CONCLUSION

In this paper, the idea of authorized data deduplication was proposed to protect the data security by including differential authority of users in the duplicate check. In public cloud our data are securely store in encrypted format,and also in private cloud our key is store with respective file.There is no need to user remember the key.So without key anyone can not access our file or data from public cloud. As a proof of concept, we implemented a prototype of our proposed authorized duplicate check scheme and conduct testbed experiments on our prototype. We showed that our authorized duplicate check scheme incurs minimal overhead compared to convergent encryption and network transfer.

REFERENCES

[1] OpenSSL Project. http://www.openssl.org/.

[2] P. Anderson and L. Zhang. Fast and secure laptop backups with encrypted de-duplication. In *Proc. of USENIX LISA*, 2010.

[3] M. Bellare, S. Keelveedhi, and T. Ristenpart. Dupless: Serveraided 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] M. Bellare, C. Namprempre, and G. Neven. Security proofs for identity-based identification and signature schemes. *J. Cryptology*, 22(1):1–61, 2009.

[6] M. Bellare and A. Palacio. Gq and schnorr identification schemes: Proofs of security against impersonation under active and concurrent attacks. In *CRYPTO*, pages 162–177, 2002.

[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] 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*, pages 617–624, 2002.

[9] D. Ferraiolo and R. Kuhn. Role-based access controls. In *15th NIST-NCSC National Computer Security Conf.*, 1992.

[10] GNU Libmicrohttpd. http://www.gnu.org/software/libmicrohttpd/.

[11] 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 491–500. ACM, 2011.

[12] J. Li, X. Chen, M. Li, J. Li, P. Lee, andW. Lou. Secure deduplication with efficient and reliable convergent key management. In *IEEE Transactions on Parallel and Distributed Systems*, 2013.

[13] libcurl. http://curl.haxx.se/libcurl/.

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

[15] W. K. Ng, Y. Wen, and H. Zhu. Private data deduplication protocols in cloud storage. In S. Ossowski and P. Lecca, editors, *Proceedings of the 27th Annual ACM Symposium on Applied Computing*, pages 441–446. ACM, 2012.

AUTHOR'S PROFILE:

SudhaRani.Chennupalli, Presently pursuing her M.Tech in Computer Science & Engineering from Chirala Engineering College, Chirala, Prakasam District, A.P, India. Affiliated to Jawaharlal Nehru Technological University, Kakinada. Approved by AICTE, New Delhi. Her B.Tech completed at Chirala Engineering College, Chirala, Prakasam District, A.P, India. Affiliated to Jawaharlal Nehru Technological University, Kakinada. A.P, India.

S. Umamaheswara Rao is an Assistant Professor in Computer Science & Engineering Department in Chirala Engineering College, Chirala, Prakasam District, A.P, India. He gained 9 years Experience on Teaching. He had Good interest on OperatingSystem, Computer Networks.