Volume 5, Issue 8, August 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Compressing the Data Secure Authorized Deduplication Checker in Hybrid Cloud

A Anusha*, Kranthi Kiran G, J Dayanika
Dept. of CSE, CMR Technical Campus,
India

Abstract: ***Data Deduplication is one of important data compression techniques for eliminating duplicate copies of repeating data and has been widely used in cloud storage in order to minimize the amount of storage space and save bandwidth. For protection of data security, this paper makes an attempt to primarily address the problem of authorized data deduplication. To protect the confidentiality of important data while supporting deduplication, the convergent encryption technique has been proposed to encrypt the data before outsourcing. Along with the data the privilege level of the user is also checked in order to assure whether he is an authorized user or not. 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 has minimal overhead compared to normal operations. As a proof of concept, we implement a prototype of our proposed authorized duplicate check scheme and conduct tested experiments using our prototype. This paper tries to minimize the data duplication that occurs in hybrid cloud storage by using various techniques.***

Keywords: Deduplication, authorized duplicate check, confidentiality, hybrid cloud, convergent encryption.

I. INTRODUCTION

Cloud computing provides seemingly unlimited "virtualized" resources to users as services across the whole Internet, while hiding platform and implementation details. Today's cloud service providers offer both highly available storage and massively parallel computing resources at relatively low costs. As cloud computing becomes prevalent, an increasing amount of data is being stored in the cloud and shared by users with specified privileges, which define the access rights of the stored data. One critical challenge of cloud storage services is the management of the ever-increasing volume of data. To make data management scalable in cloud computing, deduplication has been a well-known technique and has attracted more and more attention recently. Data deduplication is a specialized data compression technique for eliminating duplicate copies of repeating data in storage. The technique is used to improve storage utilization and can also be applied to network data transfers to reduce the number of bytes that must be sent. Instead of keeping multiple data copies with the same content, deduplication eliminates redundant data by keeping only one physical copy and referring other redundant data to that copy. Deduplication can take place at either the file level or the block level. For file level deduplication, [5] it eliminates duplicate copies of the same file. Deduplication can also take place at the block level, which eliminates duplicate blocks of data that occur in non-identical files. Cloud computing is an emerging service model that provides computation and storage resources on the Internet. One attractive functionality that cloud computing can offer is cloud storage. Individuals and enterprises are often required to remotely archive their data to avoid any information loss in case there are any hardware/software failures or unforeseen disasters.

Instead of purchasing the needed storage media to keep data backups, individuals and enterprises [6] can simply outsource their data backup services to the cloud service providers, which provide the necessary storage resources to host the data backups. While cloud storage is attractive, how to provide security guarantees for outsourced data becomes a rising concern.[8] One major security challenge is to provide the property of assured deletion, i.e., data files are permanently inaccessible upon requests of deletion. Keeping data backups permanently is undesirable, as sensitive information may be exposed in the future because of data breach or erroneous management of cloud operators. Thus, to avoid liabilities, enterprises and government agencies usually keep their backups for a finite number of years and request to delete (or destroy) the backups afterwards. For example, the US Congress is formulating the Internet Data Retention legislation in asking ISPs to retain data for two years, while in United Kingdom, companies are required to retain wages and salary records for six years. a.[7] A hybrid cloud is a combination of private cloud and public cloud in which the data which is most critical that resides on a private cloud and the data which is easily accessible is resides on a public cloud hybrid cloud is helpful for reliability, extensibility and fast deployment and cost saving of public cloud with more security with private cloud [1], [2]. The complex challenge of cloud storage or cloud computing is the arrangement of large volume of data duplication is a process of eliminating of duplicate data in de-duplication techniques redundant data removed leaving single instance of the data to be stored. In the previous old system the data is encrypted back to outsourcing[9] it on the cloud or network.

This encryption requires maximum time as well as storage space requirement to encode the data if there is large amount of data at that time encryption process becomes complex and critical. By using de-duplication technique in hybrid cloud

the encryption technique become simpler. As we all of knows that the network has large amount of data which being shared by many users. Many large networks uses data cloud to store the data and share that data on the network [3]. The node in the network have full rights to upload and download the data over the network but many time what happen the data which is being uploaded and downloaded contains same data on the network in these case data confidentiality and the security of cloud get disturbed.

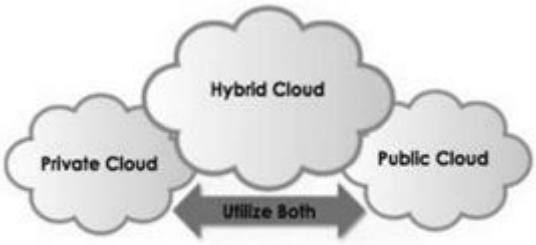

Fig. 1: Architecture of Hybrid cloud

The hybrid cloud gives the functionality, scalability, reliability, fast deployment and cost saving of public cloud storage by reducing redundancy in data [4].

II. 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[10] 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 [11]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.

System Model:

Now we see the architecture of our system. in our architecture there are three modules .
[1] User
[2] Public cloud
[3] Private cloud.etc

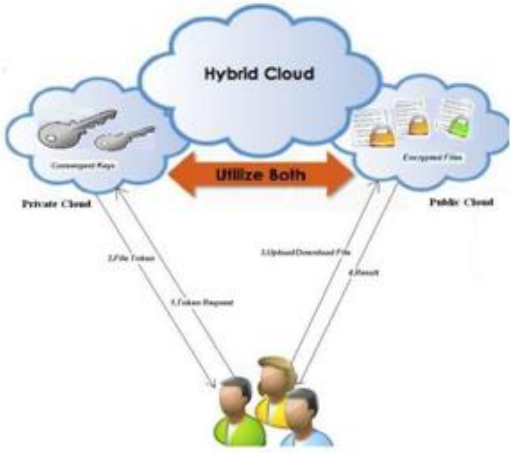

Fig 2: Architecture of Authorized Deduplication

First if the user want to upload the files on the public cloud then user first encrypt that file with the convergent key and then sends it to the public cloud at the same time user also generates the key for that file and sends that key to the private cloud for the purpose of security.In the public cloud we use one algorithm for deduplication. Which is used to avoid the duplicate copies of files which is entered in the public cloud. Hence it also minimizes the bandwidth. that means we requires the less storage space for storing the files on the public cloud.In the public cloud any person that means the unauthorized person can also access or store the data so we can conclude that in the public cloud the security is not provided. In general for providing more security user can use the private cloud instead of using [12]the public cloud.User generates the key at the time of uploading file and store it to the private cloud.When user wants to downloads the file that he/she upload,.he/she sends the request to the public cloud. Public cloud provides the list of files that are uploads the many user of the public cloud because there is no security is provided in the public cloud.When user selects one of the file from the list of files [13]then private cloud sends a message like enter the key!. User has to enter the key that he

generated for that file. When user enter the key the private cloud checks the key for that file and if the key is correct that means user is valid then private cloud give access to that user to download that file successufully. then user downloads the file from the public cloud and decrypt that file by using the same convergent key which is used at the time of encrypt that file.in this way user can make a use of the architecture.

III. RELATED WORK

However, previous deduplication systems cannot support differential authorization duplicate check, which is important in many applications. In such an authorized deduplication system, each user is issued a set of privileges during system initialization. Each file uploaded to the cloud is also bounded by a set of privileges to specify which kind of users is allowed to perform the duplicate check and access the files. Before submitting his duplicate check request for some file, the user needs to take this file and his own privileges as inputs. [14] The user is able to find a duplicate for this file if and only if there is a copy of this file and a matched privilege stored in cloud. For example, in a company, many different privileges will be assigned to employees. In order to save cost and efficiently management, the data will be moved to the storage server provider (S-CSP) in the public cloud with specified privileges and the deduplication technique will be applied to store only one copy of the same file. Because of privacy consideration, some files will be encrypted and allowed the duplicate check by employees with specified privileges to realize the access control. [15]Traditional deduplication systems based on convergent encryption, although providing confidentiality to some extent, do not support the duplicate check with differential privileges. In other words, no differential privileges have been considered in the deduplication based on convergent encryption technique. It seems to be contradicted if we want to realize both deduplication and differential authorization duplicate check at the same time.

IV. HYBRID CLOUD FORSECURE 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. In this setting, deduplication can be frequently used in these settings for data backup and disaster recovery applications while greatly reducing storage space. Such systems are widespread and are often more suitable to user file backup and synchronization applications than richer storage abstractions. There are three entities defined in our system, that is, users, private cloud and S-CSP in public cloud . The S-CSP performs deduplication by checking if the contents of two files are the same and stores only one of them. The access right to a file is defined based on a set of privileges. The exact definition of a privilege varies across applications. For example, we may define a rolebased privilege according to job positions (e.g., Director, Project Lead, and Engineer), or we may define a time based privilege that specifies a valid time period (e.g., 2014-01- 01 to 2014-01-31) within which a file can be accessed. A user, say Alice, may be assigned two privileges “Director” and “access right valid on 2014- 01-01”, so that she can access any file whose access role is “Director” and accessible time period covers 2014- 01- 01. Each privilege is represented in the form of a short message called token. Each file is associated with some file tokens, which denote the tag with specified. []A user computes and sends duplicate-check tokens to the public cloud for authorized duplicate check. Users have access to the private cloud server, a semitrusted third party which will aid in performing deduplicable encryption by generating file tokens for the requesting users. We will explain further the role of the private cloud server below. [16]Users are also provisioned with per-user encryption keys and credentials Thus, several new security notations of privacy against chosen-distribution attacks have been defined for unpredictable message. In another word, the adapted security definition guarantees that the encryptions of two unpredictable messages should be indistinguishable. Thus, the security of data in our first construction could be guaranteed under this security notion.

We discuss the confidentiality of data in our further enhanced construction.

A. Architecture For Authorized Deduplication:

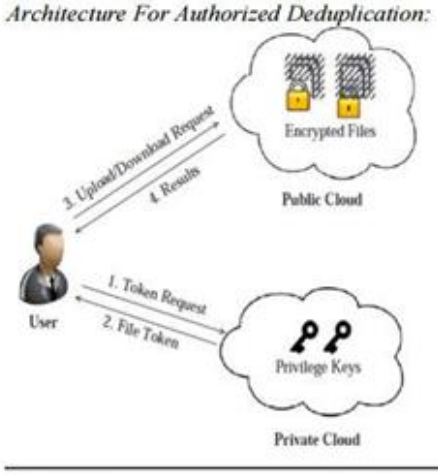

Fig Architecture for Authorized deduplication

In this paper, we will only consider the file level deduplication for simplicity. In another word, we refer a data copy to be a whole file and file-level deduplication which eliminates the storage of any redundant files. [19] Actually, blocklevel

deduplication can be easily deduced from file-level deduplication, Specifically, to upload a file, a user first performs the file-level duplicate check. If the file is a duplicate, then all its blocks must be duplicates as well; otherwise, the user further performs the block-level duplicate check and identifies the unique blocks to be uploaded. [20]Each data copy (i.e., a file or a block) is associated with a token for the duplicate check.

• S-CSP. This is an entity that provides a data storage service in public cloud. The S-CSP provides the data outsourcing service and stores data on behalf of the users. To reduce the storage cost, the S-CSP eliminates the storage of redundant data via deduplication and keeps only unique data. In this paper, we assume that S-CSP is always online and has abundant storage capacity and computation power.

• 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 user or different users.[17] In the authorized deduplication system, each user is issued a set of privileges in the setup of the system. Each 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. Specifically, since the computing resources at data user/owner side are restricted and the public cloud is not fully trusted in practice, private cloud is able to provide data user/owner with an execution environment and infrastructure working as an interface between user and the public cloud.

The private keys for the privileges are managed by the private cloud, who answers the file token requests from the users. The interface offered by the private cloud allows user to submit files and queries to be securely stored and computed respectively. [18]Notice that this is a novel architecture for data deduplication in cloud computing, which consists of a twin clouds (i.e., the public cloud and the private cloud). Actually, this hybrid cloud setting has attracted more and more attention recently. For example, an enterprise might use a public cloud service, such as Amazon S3, for archived data, but continue to maintain in-house storage for operational customer data. Alternatively, [19] the trusted private cloud could be a cluster of virtualized cryptographic co-processors, which are offered as a service by a third party and provide the necessary hardware based security features to implement a remote execution environment trusted by the users.

V. ADVANTAGES OF AUTHORISED DEDUPLICATION SYSTEM

1) The client is permitted to perform the duplicate copy check for records selected with the particular subject.
2)The complex subject to help stronger security by encoding the record with distinct privilege keys.
3)Decrease the storage space of the tags for reliability check. To strengthen the security of deduplication and ensure the data privacy.

VI. SIMULATION RESULTS

The simulation can be takes place with the help of 3 process the process one is checking the person whether he is authorized or not the data is trusted or not. And the process 2 is confidential make to keep in public or not and the process 3 is removing the duplicate and making the data unique. And again it is going to check the data having any duplicate values and that will be keep stored in the cloud and that data is encrypted with the help of convergent keys values. And also the data size will be compressed and stored in the form of blocks.

VII. CONCLUSION

Several new deduplication constructions supporting authorized duplicate check in hybrid cloud architecture, in which the duplicate-check tokens of files are generated by the private cloud server with private keys. Security analysis demonstrates that our schemes are secure in terms of insider and outsider attacks specified in the proposed security model. 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. Hybrid clouds offer a greater flexibility to businesses while offering choice in terms of keeping control and security. Hybrid clouds are usually deployed by the organizations willing to push part of their workloads to public clouds either for cloud bursting purposes or for projects requiring faster implementation Because hybrid clouds vary based on company needs and structure of implementation. In proposed system authorized data deduplication was proposed to protect the data security by including differential privileges of users in the duplicate check system presented several new deduplication constructions supporting authorized duplicate check in hybrid cloud architecture, the duplicate-check tokens of files are generated by the private cloud server with private keys. Proposed system is secure in terms of insider and outsider attacks specified in the proposed security model. The proposed authorized duplicate check scheme incurs minimal overhead compared to convergent encryption and network transfer.

REFERENCES

[1] A Hybrid Cloud Approach for Secure Authorized Deduplication Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou.

[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] 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.

[11] 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.

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

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

[14] 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.

[15] R. D. Pietro and A. Sorniotti. Boosting efficiency and security in proof of ownership for deduplication. In H. Y. Youm and Y. Won, editors, *ACM Symposium on Information, Computer and Communications Security*, pages 81–82. ACM, 2012.

[16] S. Quinlan and S. Dorward. Venti: a new approach to archival storage. In *Proc. USENIX FAST*, Jan 2002.

[17] 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.

[18] R. S. Sandhu, E. J. Coyne, H. L. Feinstein, and C. E. Youman. Role-based access control models. *IEEE Computer*, 29:38–47, Feb 1996.

[19] J. Stanek, A. Sorniotti, E.Androulaki, and L. Kencl. secure data deduplication scheme for cloud storage. In *Technical Report*, 2013.

[20] M. W. Storer, K. Greenan, D. D. E. Long, and E. L. Miller. Secure data deduplication. In *Proc. of StorageSS*, 2008.

AUTHER'S PROFILE

Anusha . A presently working as Asst. Prof. in CMR Technical Campus, Hyderabad, Telangana, India. She Received her M.Tech(CSE) from Annamarachrya Inst. Technological Sciences in 2011. She Received her B.Tech(CSE) from Sri KrishnadevarayaEngg. College in 2009 . Her area of interest is datamining, Big data

KRANTHI KIRAN G presently working as a Assistant Professor in CMR Technical campus, Hyderabad, Telangana, INDIA. He received his M.Tech (CSE) degree from Sphoorthy Engineering College in the year 2012.He received B.E. (CSE) degree from M.V.S.R Engineering College, in the year 2009. His fields of interests are Cloud Computing, Machine learning, Network Security, etc.

Dayanika J presently working as Asst. Prof. in CMR Technical Campus, Hyderabad, Telangana, India. She Received her M.Tech(CSE) from A.M Reddy Memorial College of Engg. And Tech. in 2013. She Received her B.Tech(CSE) from Nalanda Inst. Of Engg.& Tech. in 2010 . Her area of interest is datamining, Big data.

