International Journal of Engineering and Technical Research (IJETR)
ISSN: 2321-0869 (O) 2454-4698 (P), Volume-6, Issue-4, December 2016

A Secure Authorization for Cloud Data Deduplication Using Privileged Keys

Parth Walia, Khushbu Vachhani, Savi Moundekar, Rugveda Sherkar

***Abstract*— In today's world, cloud service providers offers highly handy storage and as well as bulky parallel computing re-sources at nearly low costs. As cloud computing has become frequent, an ever increasing amount of data is stored in the cloud and shared by users with particular benefits. It defines the access rights which are safeguard for stored data. One of the tough challenge of cloud storage services is the administration of data which is increasing day by day.**

For the better preservation of data security, this project makes the first pursuit for technically inscribing the question of endorsement of data deduplication. Outstanding from all the other classic deduplication systems, the differential privileges of users are further taken in application in duplicate check keeping the data separate. We also present many innovative deduplication constructions which supports authorized duplicate check in an architecture of hybrid cloud.

We are developing an advanced scheme for supporting stronger security by encryption of the file with different privilege keys. In this way, the users without correspondence of privileges do not perform the duplicate check. Security analysis shows that our system is secure in terms of the definitions which are specified in the system which is proposed.

***Index Terms*— Cloud computing, Security, Encryption Decryption Technology**

I. INTRODUCTION

Cloud computing now a day is very much used as it provides very good services. The services that the cloud is basically known for is its storage service. As the amount of data is increased so with this there arise a challenge for management of all the data that is stored in the cloud. This challenge becomes critical day by day with the increase in the data these days.

In order to make proper management of this data, cloud has a well-known technique which is known as deduplication. Nowadays this technique is gaining more attention towards itself. Date deduplication is a specialized data compression technique for eliminating duplicate copies of repeating data in storage. This technique was introduced in order to improve the storage system and proper utilization of space. When we store any copy of data on the cloud then there are chances that it might already be present over there or there are chances that someone in future might

again upload the copy of the same data. Multiple copies of same data will be there on cloud using unnecessary space repeatedly[1]. The main aim of deduplication is keep only one physical copy and remove the other copies of the same data.

Parth Walia, Student, Computer Engineering,NBNSSOE, Pune
Khushbu Vachhani, Student, Computer Engineering,NBNSSOE, Pune
Savi Moundekar, Student, Computer Engineering,NBNSSOE, Pune
Rugveda Sherkar, Student, Computer Engineering,NBNSSOE, Pune

There are two places where this method of deduplication takes place that is; at the block level or at the file level. In deduplication of block level file, it removes the duplicate blocks of data. For file level deduplication it removes the duplicate file of that data.[3]

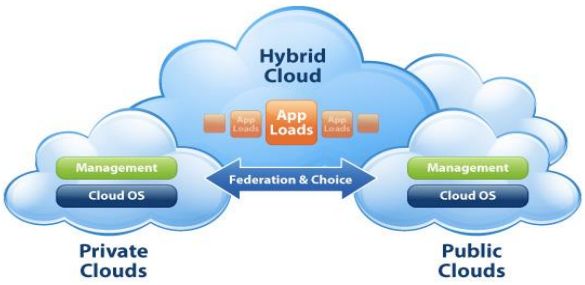
Data deduplication has advantages in form of security and privacy concerns. In Traditional system of encryption, at the time of providing data confidentially, it is incompatible with data deduplication. According to the traditional encryption technique different users require their own key in order to access the data. This will lead to the same data copies of different users making the deduplication process impossible. In this case convergent encryption makes the deduplication process feasible.[4][5] With the help of convergent key it can encrypt or decrypt a data copy. Once the key is generated and data the data in encrypted, the user can retain the key back and send ciphertext to the cloud. For the identical data copies the convergent key generated will be same and the ciphertext will also be the same. A secure proof of ownership protocol is needed in order to avoid unauthorized access if the duplicate of same file is found by the user[2][1]. Once the proof is obtained the same file will be available to the user on the server without the need to upload the same file. The encrypted file can be downloaded on the user end from a particular server and then can be decrypted but only correlative data owner's confluent keys. So in this way convergent keys allow the deduplication with the ownership proof.

II. LITERATURE SURVEY

In previous deduplication system there were certain issues as for same data multiple copies were created on the cloud leading to unnecessary wastage of storage space. The goal here is to minimize the duplicate data on the cloud and even increase the security level. It is an advanced theory for backing stronger security by encoding a file with different privileged keys. In this way the user cannot perform duplicate check. With these the integrity of the system is improved. Storage utilization in increased. Once the duplicate copies are eliminated it will improve the reliability.

III. OVERVIEW OF THE HYBRID CLOUD CONCEPTS

From several years, the agenda of cloud computing has been the focus of IT and corporate world, but the extremely conscious one with the security matter have been hesitant to move their workloads and information into the cloud. Now, with the existing technology within cloud services available for deployment in organizations, a versatile model of cloud computing is rapidly gaining a foothold in business: The Hybrid Cloud.



[1]The duplication of data and to maintain the confidentiality in the cloud is a major issue nowadays so to handle this we use the concept of Hybrid Cloud. It is a union of private and public cloud. Hybrid cloud storage provides a variety of advantages such as flexibility, reliability, rapid implementation and potentially reducing cost of public cloud storage with the surveillance and full control over private cloud storage[6].

Usage of a hybrid cloud can incredibly ease accordance in the workplace. Public cloud beneficence isolate do not easily integrate with on-premises hardware. Devices such as printers, scanners, fax machines, and physical security hardware, like security cameras, fire, and CO_2 detectors, can be burden to public cloud acceptance[6]. Rather than leaving the following mission-critical devices from the rest of the organization's network, the use of a private cloud element would be highly profitable.
Combination of public services and private clouds with the data center resembling hybrid is the new era for corporate computing. This is not compulsory that all companies using some public and some private cloud services and tend to have a hybrid cloud. Rather, a hybrid cloud is a platform where the public and private services are used together to create profit[7].

A cloud is hybrid: If a company utilizes a civic progress platform that dispatches data to a personal cloud or a facts hub–based applicances.

When a company drags a number of SaaS (Software as a Service) applications and shifts data between data center resources or private.

A cloud is not hybrid: If some developers in a company use a public cloud service to predate a new app that is totally detached from the data center or private cloud.

If SaaS application is used by a company for a project but there is no exchange of data from that application into the company's private cloud.

The ratification of hybrid cloud models is anticipated to grow at such high rates that nearly half of the larger organizations would have adopted hybrid platforms in working by 2017, A hybrid model provides services with a competitive edge through on-demand resource utilization, the capability to swap between clouds, and elasticity with high-end solutions such as: Cloud integration, Cloud management, Cloud security, Automation, Networking, Consulting[7].

The major hybrid cloud providers are: VMware, Amazon Web Services, Microsoft, Rackspace, EMC Corp, Hewlett Packard, IBM, Cisco system, Dell, Eucalyptus.

IV. SECURE DEDUPLICATION SYSTEMS

Secure Data deduplication is a data compression method used for removing the duplicate data. It is commonly present in cloud storage to minimize the storage space and also to save bandwidth. This technique is proposed to keep the confidentiality of sensitive data. This is used for making the feasible deduplication and maintain data confidentiality[7]. Hybrid Cloud architecture contains many innovative deduplication constructions which supports authorized duplicate check[8]. Proposed security models contain demonstration of security analysis scheme. Cloud storage service is the mechanism for evergreen increasing mass of the data. This technique which is used for improving the storage utilization. It is also used as an application for network data transfers to reduce number of bytes that are to be sent. Data deduplication occurs in two ways-block level as well as file level. The duplicate copies of identical file eliminate by file level deduplication. [8][9]Data deduplication takes a lot of benefits. Security as well as privacy issues arise as user's sensitive data are capable to both insider and outsider attacks. Traditional encryption requires different users to encrypt their data with own keys[9]. There are two techniques used for secure data deduplication: 1) Key Management 2) Convergent Encryption .It provides virtual infrastructure to host application services. These services can be used by client to manage his data stored.

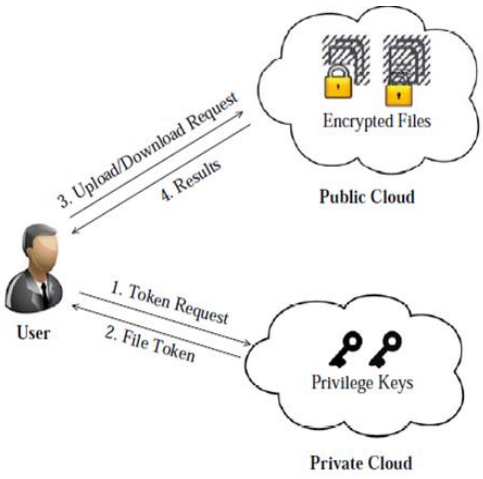


Fig. Architecture for Authorized Deduplication

V. OUR PROPOSED SYSTEM

In the existing Data deduplication systems, to grant data owner to firmly perform check on duplication with differential benefits the private cloud is intrigate as a standard. The architecture like this is reasonable and has claimed heavy attention of researchers. The data owners only out-source their data storage by exploiting public cloud.At the same time, the operation of the data is manageable in cloud which is private.

The disadvantages of the existing system is, of the Traditional encryption, while providing data confidentiality, and is incompatible with data deduplication. Same data copies, but of different users will obtain different ciphertext, which makes the deduplication impossible.

In the proposed system we are achieving the data deduplication by providing the proof of data by the data owner. This proof is used at the time of uploading of the file. Each file which is uploaded to the cloud is also bounded by a set of privileges which specifies the kind of users which allows to perform the duplicate check and helps to access the files[1]. Before submission of the duplicate check request for some file, it is necessary for the user to take this file and inputs his own previleges. The user finds a duplicate for this file but only if there is a copy of this file and privileges which matches the data stored in cloud.

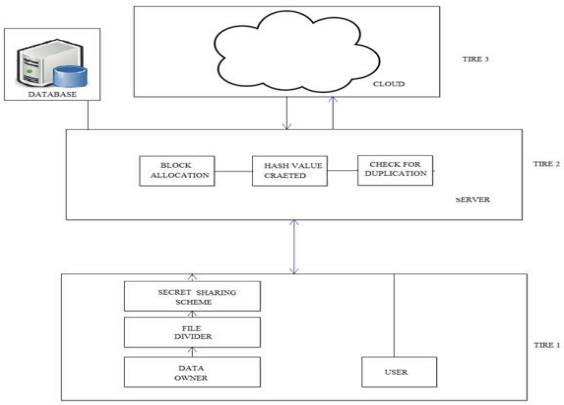


Fig. System Architecture

Cloud Service Provider

In this module, we will be developing Cloud Service Provider module. Which 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 unnecessary data via deduplication and keeps only exclusive data[10]. In given paper, we estimate that S-CSP is always online and has ample storage capacity and calculate power.

Data Users Module

A user is subsistence that expands data storage to the S-CSP and balances the data later. In a storage system backing up deduplication, the user only sends exclusive data but does not upload any duplicate data to save the uploaded bandwidth, which might be retained by the same user or different users[11]. In the authorized deduplication system, each user is delivered a set of allowance in the setup of the system. Each file is secured with the convergent encryption key and privilege keys to understand the authorized deduplication with differential privileges[12].

Private Cloud Module

Correlated with the conventional deduplication planning in cloud computing, this is a new article suggested for ease of user's secure usage of cloud service.

Precisely, since summing assets at data user/owner side are limited and the public cloud is not entirely creditable in perform, private cloud is able to supply data user/owner with an implementation atmosphere and transportation functioning as an edge between user and the public cloud[1][2][3]. The confidential keys for the rights are handled by the private cloud, who comments 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.

Secure Deduplication System

We consider several types of privacy we need protect, that is, un-forget ability of duplicate-check token: There are two variety of adversaries, that is, external adversary and internal adversary. The exterior adversary can be viewed as an internal adversary exclusive of any privilege.

If a client has opportunity p, it requires that the adversary cannot form and output a valid duplicate token with any other privilege p′ on any file F, where p does not match p′. In addition, it also requires that if the adversary does not make a demand of token with its possess privilege from private cloud server, it cannot fake and yield a valid duplicate token with p on any F that has been queried.

VI. ALGORITHMS USED

1) CONVERGENT ENCRYPTION

Convergent encryption produce data affinity in de-duplication. A user (or data owner) acquires a convergent key from each unique data duplicate and encodes the data copy with the convergent key. In summing up, the user also obtains a tag for the data copy, such that the tag will be used to perceive duplicate[1]. Here, we suppose that the tag accuracy assets holds, i.e., if two data copies are similar, then their tags are the similar. To spot duplicate, the user first sends the tag to the server side to make sure if the alike copy has been previously stored[4]. Make a note of that both the convergent key and the tag are separately derived and the tag cannot be used to figure out the convergent key and negotiate data secrecy. Mutually the encrypted data copy and its equivalent tag will be stored on the server side.

A convergent encryption scheme can be defined with four primitive functions:

1. **KeyGenCE (M)! K** is the key production method that outline a data duplicate M to a convergent key K.
2. **EncCE (K, M)! C** is the proportional encryption method that takes mutually the convergent key K and the data replica M as inputs and then outputs a ciphertext C;
3. **DecCE(K, C)!M** is the decryption algorithm that takes in cooperation the ciphertextC and the convergent key K as inputs and then outputs the inventive data copy M; and
4. **TagGen(M)! T (M)** is the tag production method that converges the creative data duplicate M and outputs a tag T (M).

2) Proof of Ownership

POW which is noted as a proof of ownership is the one due to which a user enables the proof of ownership to the storage server[1][2].

Proof of ownership is implemented as interactive algorithm. Short value (M) is derived from a data copy M, there are certain steps to be carried out by the verifier M.[1]

Pseudo code

Step 1: Calculate the two convergent key values
Step 2: Compare the two keys and files get accessed.
Step 3: Apply de-duplication to eradicate the duplicate values.
Step 4: If any other than the duplicates it will be checked once again and make the data unique.
Step 5: That data will be unique and also more confidential the authorized can access and data is stored.

VII. FUTURE SCOPE

The security problems that may arise in the present practical model can be excluded. By deduplicating the data, the memory wastage is minimized. It provides authorization to private firms and protect the confidentiality of the important data.

REFERENCES

[1] J. Li, X. Chen, M. Li, J. Li, P. Lee, andW. Lou. Se-cure deduplication with efficient and reliable convergent key management. In IEEE Transactions on Parallel and Distributed Systems, 2013.
[2] P. Anderson and L. Zhang. Fast and secure laptop back-ups with encrypted de-duplication. In Proc. of USENIX LISA, 2010.
[3] C. Ng and P. Lee. Revdedup: A reverse deduplication storage system optimized for reads to latest backups. In Proc. of APSYS, Apr 2013.
[4] M. Bellare, S. Keelveedhi, and T. Ristenpart. Dup-less: Serveraided encryption for deduplicated storage. InUSENIX Security Symposium, 2013.
[5] W. K. Ng, Y. Wen, and H. Zhu. Private data dedu-plication 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.
[6] M. Bellare, S. Keelveedhi, and T. Ristenpart. Mes-sage-locked encryption and secure deduplication. In EU-ROCRYPT, pages 296–312, 2013.
[7] M. Bellare, C. Namprempre, and G. Neven. Security proofs for identity-based identification and signature schemes. J. Cryptology, 22(1):1–61, 2009.
[8] 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.
[9] 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.
[10] 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.
[11] D. Ferraiolo and R. Kuhn. Role-based access controls. In 15th NIST-NCSC National Computer Security Conf.,1992.
[12] 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.
[13] S. Quinlan and S. Dorward. Venti: a new approach to archival storage. In Proc. USENIX FAST, Jan 2002.

Parth Walia is Student at NBNSSOE, Pune. He is Currently pursuing his BE degree in Computer Engineering. Parth has vast knowledge on the cloud computing and its applications, and shares an interest towards the development of this field.

Khushbu Vachhani is Student at NBNSSOE, Pune. She is Currently pursuing his BE degree in Computer Engineering. Khushbu has vast knowledge on the cloud computing and its applications, and is doing many studies on the field.

Rugveda Sherkar is Student at NBNSSOE, Pune. She is Currently pursuing his BE degree in Computer Engineering. Rugveda has very good knowledge on the cloud computing and its applications, and development.

Savi Moundekar is Student at NBNSSOE, Pune. She is Currently pursuing his BE degree in Computer Engineering. Savi has very well knowledge on the domain of cloud computing, its applications, and development