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. 3, Issue 11, November 2015

A Study on Composite Approach for Availing Deduplication in a Secure and Authorize Paradigm

Shabnam Siddiqui[1], Amit Zore[2]

M.E Student, Dept. of Computer Engineering, DPCOE, Savitribai Phule Pune University, Pune, India[1]

Assistant Professor, Dept. of Computer Engineering, DPCOE, Savitribai Phule Pune University, Pune, India[2]

ABSTRACT: Today, use of cloud computing is increasing rapidly. Cloud computing is very important in the data sharing application. Daily use of cloud is increasing. But the problem in cloud computing is every day knowingly or unknowingly, increasing similar data copies are uploaded on the cloud. Therefore we can reduce the size of redundant or duplicate data in cloud using the data DE duplication method. This method main aim is that remove duplicate data from cloud. It can also help to save storage space and bandwidth. Our proposed method is to remove the duplicate data but in which user have assigned some privilege according to that duplication check. Cloud DE duplication is achieve using the hybrid cloud architecture. We proposed method is more secure and consumes less resources of cloud. Also we have shown that proposed scheme has minimal overhead in duplicate removal as compared to the normal DE duplication technique.

KEYWORDS: Authorization; data security; privilege; DE duplication; credentials; cloud.

I. INTRODUCTION

Current era is cloud computing era. Now a days cloud computing has wide range of scope in data sharing. Cloud computing is provide large amount of virtual environment hiding the platform and operating systems of the user. User uses the resources for sharing data. But users have to pay as per the use of resources of cloud. Now cloud service providers are offering cloud services with very low cost and also with high reliability. User can upload the large amount data on cloud and shared data to millions of users. Cloud providers offer different services such as infrastructure as a service, platform as a service, etc on rental basis. User not needs to purchase the resources. As the data is get uploaded by the user every day it is critical task to manage this ever increasing data on the cloud. DE duplication is best method to make well data management in the cloud computing. This method is becoming more attraction for data DE duplication.

This method sends the data over the network required small amount of data. These methods have application in data management and networking. Data duplication is the technique of reducing the size of data Also it is the best compression method for the data DE duplication. This method is sending the data over the network required small amount of data. These methods have application in data management and networking. Instead of keeping redundant copies of the same data DE duplication only keep original copy and provide only references of the original copy to the redundant data. There are two methods of the duplication check, one is file level duplication check and other is block content level duplication check. In the file level duplication check is remove the same name file from the storage and block level DE duplication are removed the duplicate blocks. As the data DE duplication is considering the user data there must be need of the some security mechanism. It arises security and privacy concern of the user's sensitive data. In the traditional method user need to encrypt his own data by himself so there are different cipher files for each new user.

DOI: 10.15680/IJIRCCE.2015. 0311126

To avoid the unauthorized data DE duplication convergent data DE duplication is proposed to enforce the data

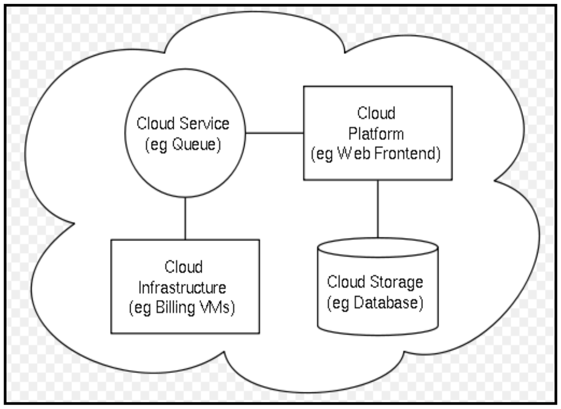


Fig 1. Cloud architecture and services

confidentiality while checking the data duplication. The clouds servers users with many services as shown in the below figure such as platform, services, infrastructure as a service, and database as a service. In this we are using thecloud storage as a service. We are using user credentials to check the authentication of the user. In the hybrid cloud there is a combination of two types of clouds namely, private cloud and public cloud. In private cloud store the user credential and the user data present in public cloud. The hybrid cloud take advantages of both public cloud and private cloud as shown in the figure 2. public cloud and private cloud are present in the hybrid cloud architecture When any user forward request to the public cloud to access the data user needs to submit his information to the private cloud then private cloud will provide a file token and user can get the access to the file which resides on the public cloud. This techniqueuses a hybrid cloud architecture in proposed. The file name is check on primary level in file data duplication and data DE duplication is checked at the block level. If user wants to retrieve his data or download the data file, he need to download both the files from the cloud server this will lead to perform the multiple operations on the same file and this violates the security of the cloud storage

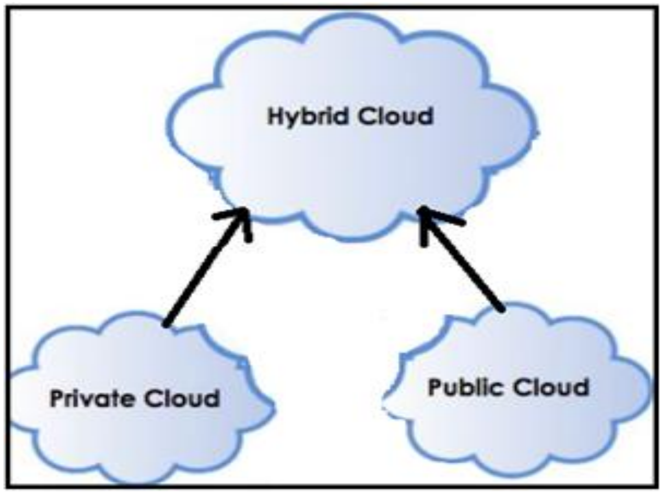


Fig 2. Hybrid Cloud Architecture

II. RELATED WORK

Cloud computing has emerged as an interestingly new technology which aids in "virtualized" Resources for all internet or cloud users as service throughout the Internet. Now a day's CSP’s other are loaded with freely facilitated storage and enormously distributed computing resources at comparatively low costs. As cloud computing users increases and cloud usage became common, a large amount of data and files are being uploaded over the clouds and are also being shared amongst different cloud users with distinguished privileges, which maintain the access rights or access policy for files on the storage. The critical challenge on management of data over clouds is to manage the continuously increasing data storage over clouds. So to make this data management on storages over clouds convenient, deduplication of data has emerged as an area of interest so as to reduce redundant data being stored over clouds and thereby make clouds free from duplicate data. Data deduplication is a novel data compression methodology which avoids maintaining duplicate copies of the data being stored on clouds unnecessarily. This methodology helps in improving the storage space optimal utilization and it can also be used in network usage by keeping the track of byes being sent over the network and avoiding sending of duplicate bytes over the same network. Instead of maintaining the duplicate data copies, this technique will make the use of only single storage space for one data copy and other data can be addressed from the data copy being stored. Deduplication process can be carried out on two mechanisms: file level deduplication which cares only for the file names and block level deduplication which works for file contents. In file level deduplication, duplicate named files are avoided from being stored on clouds. And block level deduplication deals with the contents in the files which checks the data block from the file and avoids duplication of identical data over the clouds.

Hence, But identical data copies of different users is not considered as the same data copy as individual user is allocated with its individual storage spaces over the clouds, making deduplication impossible. Hence convergent encryption aids cloud providers in obtaining or making deduplication possible. It encrypts or decrypts the data fileusing a convergent key that is gained by getting the cryptographic hash value of the datain the file.

Post key generation and post data encryption, users can get the keys and send the encrypted data over the cloud storages. As encryption operation is predefined and uses similar keys, identical data files will generate the same hash values and same convergent keys. To avoid the unauthorised access over cloud storage, a novel proof of ownership mechanismis required so as to assure that it’s the same user who owns the identical file found over the cloud. Post submission of POW, cloud user can download and decrypt the data from the cloud storage.

III. LITERATURE SURVEY

1. DupLESS: Server-Aided Encryption for Deduplicated Storage

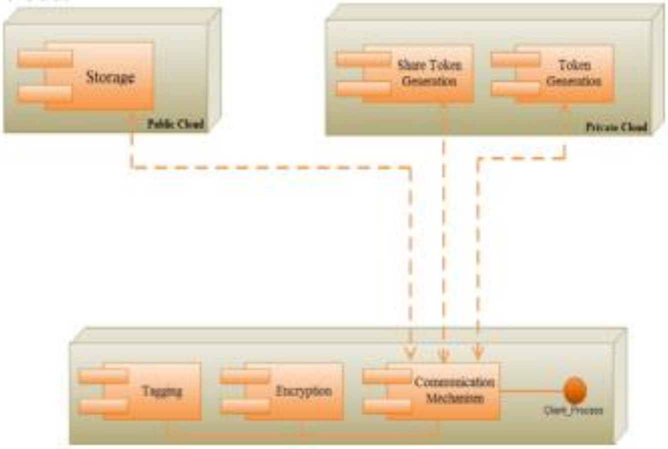
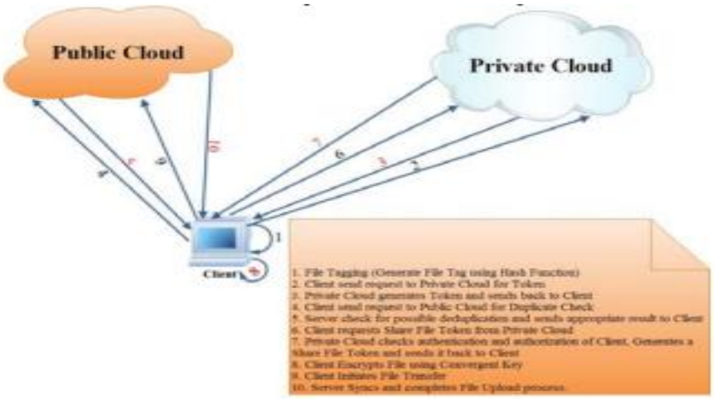


Fig3. System Framework Communication

DOI: 10.15680/IJIRCCE.2015. 0311126

For addressing the problem of authorized data deduplication the first attempt was made in this system. From the traditional system the deduplication system is different. In the proposed security modelin terms of the definitions specified this scheme is secure and demonstrates security analysis. The proposed system for authorized duplicate checkscheme is shown in this system compared to normal operations which incurs minimal overhead and therefore can achieve higher performance by encryption for deduplicated storage [1].



1. File Tagging (Generate File Tag using Hash Function)
2. Client send request to Private Cloud for Token
3. Private Cloud generates Token and sends back to Client
4. Client send request to Public Cloud for Duplicate Check
5. Server check for possible deduplication and sends appropriate result to Client
6. Client requests Share File Token from Private Cloud
7. Private Cloud checks authentication and authorization of Client, Generates a Share File Token and sends it back to Client
8. Client Encrypts File using Convergent Key
9. Client Initiates File Transfer
10. Server Syncs and completes File Upload process

Fig 4.Proposed System of Operations

2. Fast and Secure Laptop Backups with Encrypted De-duplication

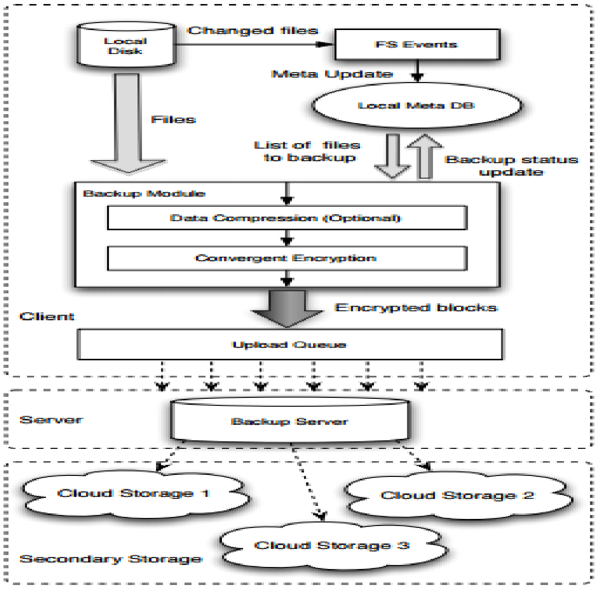


Fig 5: Architecture of the proposed backup system

For increasing the speed of backups, and reduce the storage requirementsthis system describes an algorithm. The advantage of the data is takenin this algorithm which is common between users. The per-user encryption is supported by this algorithm for maintaining confidentialityof personal data, which is necessary. [2].

3. Secure deduplication with efficient and reliable convergent key management.

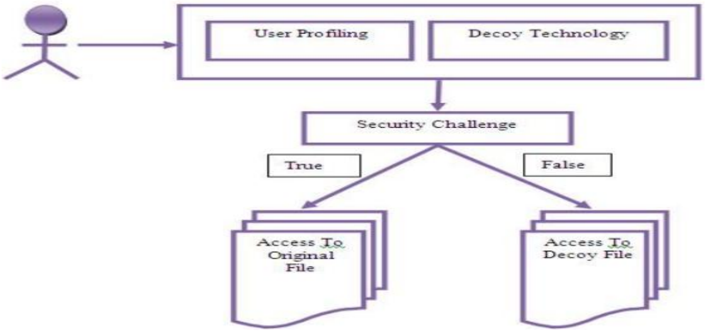


Fig 6: Outsider Attacker data security

In this system, Decoys technology, User Behavior Profiling and Dekey are proposed. Acrossmultiple servers for insider attacker there is no need for user to manage any keys on their own but instead securely distribute the convergent key shares. Using the Ramp secret sharing scheme, this system implements Dekey as a proof of concept anddemonstrates that in realistic environments. Limited overhead is incurs by Dekey.

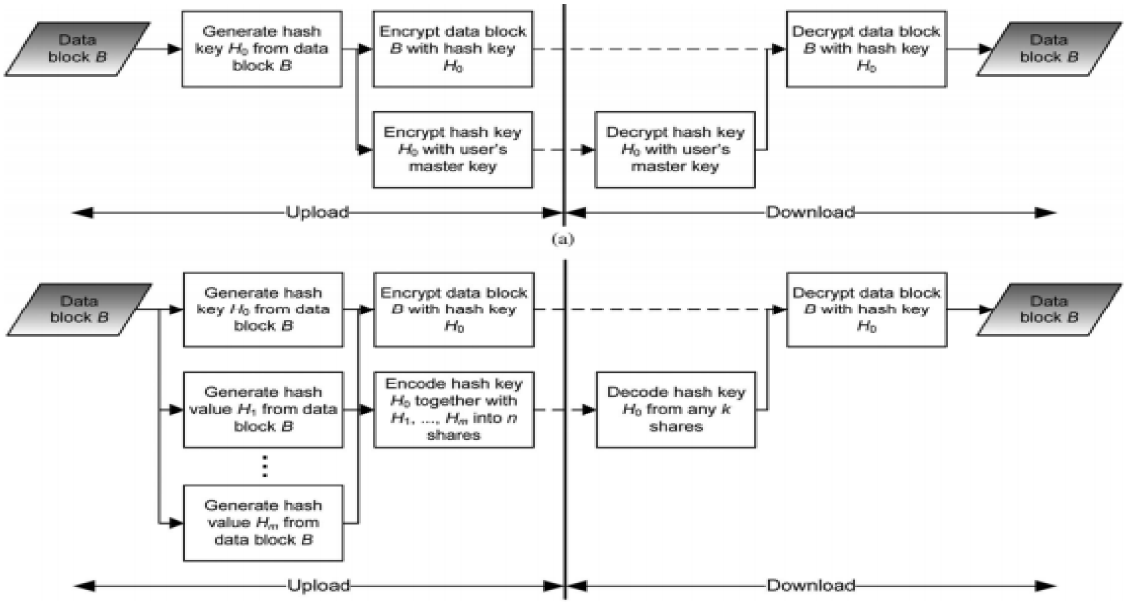

Data block B
Generate hash key H_0 from data block B
Encrypt data block B with hash key H_0
Encrypt hash key H_0 with user's master key
Decrypt hash key H_0 with user's master key
Decrypt data block B with hash key H_0
Data block B
Upload
Download
(a)
Data block B
Generate hash key H_0 from data block B
Generate hash value H_1 from data block B
Generate hash value H_m from data block B
Encrypt data block B with hash key H_0
Encode hash key H_0 together with H_1, ..., H_m into n shares
Decode hash key H_0 from any k shares
Decrypt data block B with hash key H_0
Data block B
Upload
Download

Fig7: Secure deduplication

Then, for two purposes, user profiling and decoys are served:First one is whether data access is authorized it is validating when there is detectedabnormal information access, and second one is thatwith bogus informationconfusing the attacker. We posit that for the deduplication in insider as well as outsider attacker, tis system will provideunprecedented levels of security with the combination of these security features [3].

4. A reverse deduplication storage system optimized for latest backups
5.

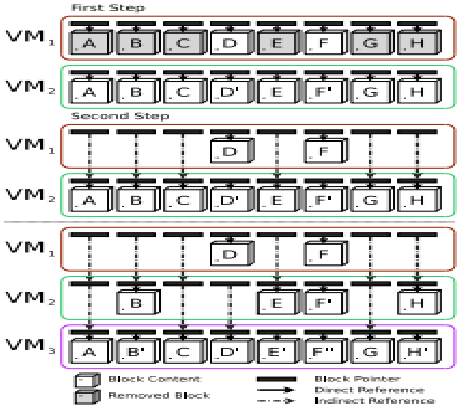


Fig 8: An example of reverse deduplication for multiple versions of the same VM.

Effectively eliminates duplicates on cloud storages. Using reverse deduplication, for optimizing leads to the latest backups of virtualmachine (VM) imagesRevDedup, a deduplication system proposed in this system. From new dataduplicates are removed in contrast with conventional deduplication, from old dataRevDedup, which removes duplicates, while keeping the new data'slayout as sequential aspossiblethereby shifting fragmentation to old data [4].

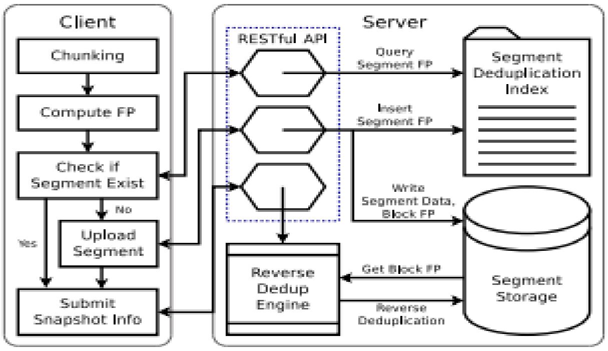


Fig 9: RevDedup's client-server model.

6. Twin Clouds: An Architecture for Secure Cloud Computing

To an untrusted commodity cloud for secure outsourcing of data as well as arbitrary computations architectureis proposedin this paper. In our approach,in the untrusted commodity cloud,for encrypting as well as verifying the data stored and also performed operations, the user communicates with a trusted cloud. In the less time-critical setup phase,for security-criticaloperationsthe computations such that the trusted cloud is mostly used are splits in this paper, whereas on encrypted data, by the fast commodity cloud, there are processed in parallel the queries to the outsourced data [5].

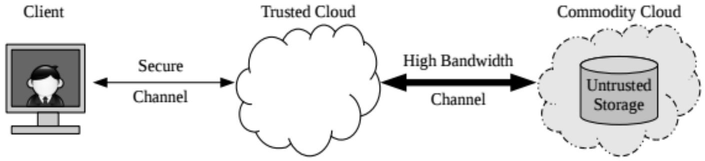


Fig. 10.System Model

7. A secure cloud backup system with assured deletion and version control

In this paper, with the policy based file access by providing access to the files, we implemented secure cloud storage. The combination of the user's credentials is the Private Key. So that there will be achieved high security.

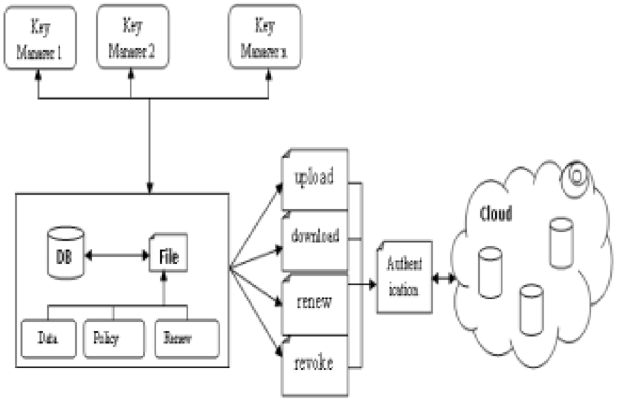


Fig 11: Overall system diagram

For file assured deletion there scheme is used the time based file Revocation. There will be automatically revoked the file when the time limit of the file expired, and in future it cannot be accessible to anyone. There also supported the Manual Revocation. This paper proposed the Policy based file renewal. By providing the new key to the existing fileThe Renewal can be done, and until the new time limit reacheswill remains the file [6].

IV. CONCLUSION

Here we can conclude that our proposed system data DE duplication of file is done authorizes way and securely. In this we have also proposed new duplication check method which generate the token for the private file. The data user need tosubmit the privilege along with the convergent key as a proof of ownership. We have solved more critical part of the cloud data storage which is only tolerated by different methods. Proposed methods ensures the data duplication securely.

REFERENCES

[1]. M. Bellare, S. Keelveedhi, and T. Ristenpart. Dupless: Serveraided encryption for deduplicated storage. In USENIX Security Symposium, 2013.
[2]. P. Anderson and L. Zhang. Fast and secure laptop backups with encrypted de-duplication. In Proc. of USENIX LISA, 2010.
[3]. 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.
[4]. C. Ng and P. Lee. Revdedup: A reverse deduplication storage system optimized for reads to latest backups. In Proc. of APSYS, Apr 2013.
[5]. 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.
[6]. 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.