Private and Public Cloud Based Security Enabled Approach for Safe Authorized Elimination of Duplicate Copies of Files

Sharana Basappa M V
Department of Computer Science and Engineering,
Dr.Ambedkar Institute of Technology,
Bangalore-560056, Karnataka, India

M V Vijaya Kumar
Department of Computer Science and Engineering,
Dr. Ambedkar Institute of Technology,
Bangalore-560056, Karnataka, India.

***Abstract*—In the area of Cloud computing, the word Cloud is used as metaphor for "the internet". The phrase cloud computing means "a type of internet based computing where different services such as servers, storage and applications are delivered to an organizations. In this paper, we deal with the problem of duplication of files. Security is also provided to the original file by the way of encryption. Eliminating duplication of files helpes in reducing the storage space and saves bandwidth during file transfer, to protect the sensitive data.The encryption technique is used for encrypt the data before storing in cloud. In this blueprint there is only a single copy of original file which is retained while others are deleted and that original file is encrypted and stored in the cloud. Conflient encryption is used. A user (data owner) derives a confluent key from each original data copy and encrypts the data copy with confluent key. In addition, the user also derives a logo for data copy, such that logo will be used to detect duplicates. If two data copies are same, then their logos are the same. To find duple, the user initially sends the logo to the server side to check if the identical copy has been already stored. Both the encrypted data copy and its correlative logo stored on server side. File-Logo (file) it computes SHA-1 hash of the file as file-logo.**

Keywords—Cloud, duplication, Bandwidth, Encryption, Data owner, Confluent key, Flie-Logo, Security, SHA-1.

I. INTRODUCTION

Cloud technology has gradually become visible and is a very important aspect in many businesses arenas. In public cloud to store the data or information, in terms of unoccupied space and time is full of energy for backup and recovery, social network systems etc. This paper focus on the basic structure services for the society [1], some of the useful data optimization services that the cloud offers. This architecture benefits both cloud services providers and cloud users, these amount of space can be saved in the cloud having a single copy of the file [2], this result in reduction of the cost [3] and increase in storage efficiency, while improving user experiences, duplication can takes place at either file level or block level. For file level deduplication it discards the duplicate copies of the same file. In block level which eliminates duplicate block of data [4] the concept of sensitive data stored in the cloud in encrypted manner, to protect the sensitive data

stored in the cloud against the disclosure by satisfying some of secrete requirements or information [5] to eliminating duplicate copies of the files bring a lot of benefits, security and privacy protection that is concerns arise as users sensitive data are harmed by particular thing to both insiders and outsiders attacks. The user is able to find the duplicate file at the time of uploading the files to the cloud. Supporting duplicate check with differential privileges, if we want to realize both deduplication and differential authorization duplicate check at the same time.

II. LITERATURE SURVEY

A. detailed description of cloud computing.

Cloud computing is the carting of computing services over the Internet. Cloud services grantunit and businesses to use software and hardware that are managed by third parties at remote locations. For instance of cloud services include online file larder, social networking layout, e-mail, and online business function. The cloud computing model allows access to information and computer resources from anywhere that a network connection is available. Cloud computing provides a mutual tarn of revenue, including data storage area, networks, computer processing bent, and particular corporate and user applications.

B. Deduplicatin and compression technique in cloud design

Our approach to deduplication results in reduction ofstorage area and transmission capacity usage during file transfers. The blueprint depends on multiple metadata structures for deduplication. [7] Only single a copy of theduplicate files is retained while others are discards. The reality of duplicate files is decisive from the metadata. The files aredivided into chunks based on their size. They are then chunk parts,de-duplicated, reducedand saved. Chunking restrictsthe number of segments and their sizes so that it is optimum for each file capacity. When the user requests a file, compressed segmentsof the file are sent over the network along with the file-to-chunkmapping.These are thenot-compressed and joined to create a complete file, hence minimizing transmission capacity functions.

C. A Secure Data Deduplication Schemefor Cloud Storage

Data according to their trade.Based on this plan, we blueprint an encryption method that guarantees semantic

ISSN: 2278 – 7798

security for drip data and provides less security and better storage and transmission capacity benefits for trendy data. This way, data deduplication can be effective for trendy data, whilst semantically secure encryption protects drip content. We show that our method is secure under the Symmetric External As more corporate and private users outsource their datato cloud storage providers, recent data breach incidents make end-to-end encryption an increasingly extrusive requirement. Dismally, semantically secure encryption schemes render various cost-effective storage optimization methods, such as data deduplication, feeble. [8]The authors present a novel idea that differentiates Decisional diffie-Hellman Assumption in the random oracle model.

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

Many people now store large amount of useful and corporate data on laptops, computers etc. These oftenhave poor or intermittent connectivity, and are weakto crime or hardware breakdown. Current backupsolutions are not well suited to this environment, andbackup systems are frequently incompetent. [9].This paperspecifies an algorithm which takes advantage of the datawhich is common between users to increase the speed ofbackups, and reduce the storage functions. This algorithmsupports client-end per-user encryption which isnecessary for personal data. It also supportsa single feature which allows fast detection ofcommon sub-trees, avoiding the need to query the backupsystem for every file. We describe a model implementationof this algorithm, and present ananalysis of the potential effectiveness, using real data obtainedfrom a set of typical users. Finally, we discuss theuse of this model in agreement with remote cloudstorage, and present an analysis of the typical price savings.

E. A Secure Cloud Backup System with Assured Deletion and Version Control.

Cloud storage is egress utility standards that empower original and enterprises to deploythe storage ofinformation backups to remote cloud providers at a low price. However,cloud users must incite security assurance of their deployeddata necessity. We present blanch-version, a secure cloud backupsystem that serves as a security layer on top of today's cloud storageservices. Blanch-Version follows the standard version-controlledbackup design, which avoids the storage of surplusdataacross different versions of necessity. On top of this, Blanch-Versionapplies cryptographic safety to data backups. Specifically, itenables fine-grained assured identification, that is, cloud users canassuredly discards particular useful versions or files on the cloudand make them permanently far to anyone, while otherversions that share the common data of the deleted versions orfiles will remain unaffected. [10] We implement a clue-of-conceptprototype of BlanchVersion and conduct factual evaluation atopAmazon S4. We show that BlanchVersion only adds lowperformance aerial over a traditional cloud backup servicethat does not support sealed deletion.

F. Security Proofs for Identity-Based Identification and Signature Schemes.

This paper provides the elimination of duplication in the cloud and reduced storage capacity. Uses AES encryption method to provide security for the file stored in the cloud server. Either security clue or blitzkrieg for a big numberof signature methods defined either explicitly or implicitly in existing literature [11]. Underlying these are a framework that onthe one hand helps explain how these methods are copied, and on the other handEmpowerthe elimination of reduplication.

G. Twin Clouds: An Architecture for Secure Cloud Computing.

Cloud computing guarantees a more price effective permissive methods to outsource storageand computations. Existing approaches for secure outsourcing of information and arbitrary computations areeither based on a individual tamper-proof hardware, or based on recently proposed fully Homomorphic encryption.[12]The hardware based solutions are un-scalable, and fully homomorphic encryption iscurrently only of theoretical interest and very in-efficient.In this paper we propose architecture for safe outsourcing of information and erratic computations toan untrusted commodity cloud. In our approach, the user communicates with a devoted cloud (eithera private cloud or built from multiple secure hardware modules) which encrypts and varies the datastored and operations performed in the un-devoted commodity cloud. We divide the computations suchthat the devoted cloud is mostly used for safe-critical functions in the low duration-critical setupstage, whereas queries to the outsourced data are processed in parallel by the fast commodity cloudon encrypted data.

H. A Hybrid Cloud Approach for Secure Authorized Deduplication

Data triplication is one of vital data confining method for delete duplex copies of same file,the definitions specified in the actualizesecurity stage [13]. As a clue of concept, we actualizeaideal of our proposed authorized duplicate check method and conductfling-bed experiments using our paradigm. We show that our proposed authorized duplex check method incurs less overheadcompared to regular operations.And has been widely used in cloud storage to reduce the amount of storage space and save transmission capacity. To protect the confidentialityof sensitive data while supporting deduplication, the concurrent encryption method has been introduced to encrypt the data beforesending. To better protect data security, this paper makes the initial attempt to formally point to the trouble of authorized datatriplication. Different from natural triplication systems, the rackprivileges of users are further considered in duplicatecheck besides the data itself.We also present several new deduplication constructions supporting authorized duple check in thePublic and Private cloud construction.

ISSN: 2278 – 7798

III. PROBLEM STATEMENT

The problem of the system is to private cloud which is involved as a proxy to allow data owner/users to securely perform duplicate check with differential privileges. Such architecture is practical and has attracted much attention from researchers. The data owners only outsource their data storage by utilizing public cloud while the data operation is managed in private cloud.

IV. USE-CASE

The data owner uploads their data in the cloud server. For the security purpose the data owner encrypts the data file and then store in the cloud. The data owner can check the duplication of the file over Corresponding cloud server. Data usercan only access the data file with the encrypted key if the user has the privilege to access the file. For the user level, all the privileges are given by the Domain authority and the Data users are controlled by the Domain Authority only.

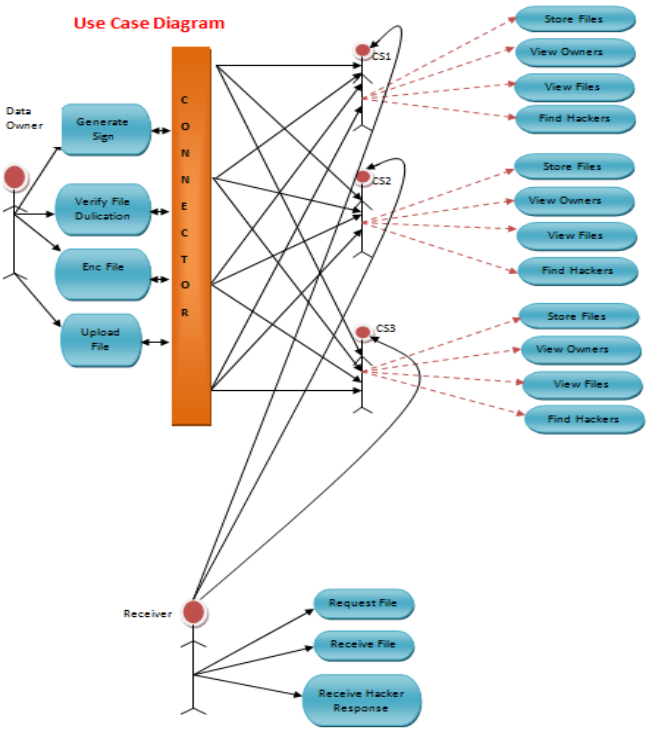

Fig 1. Use Case Diagram

The cloud service provider manages a cloud to provide data storage service. Data owners encrypt their data files and store them in the cloud for sharing with data consumers. To access the shared data files, data consumers download encrypted data files of their interest from the cloud and then decrypt them.

A user is an entity that wants to outsource data storage to the CSP and access the data later. In a storage system supporting re-duplication, 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. In the authorized de duplication 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 de duplication with differential privileges.All the legal users in the system can freely query any interested encrypted and decrypted data. Upon receiving the data from the server, the user runs the decryption algorithm Decrypt to decrypt the cipher text by using its secret keys from different Users. Only the attributes the user possesses satisfy the access structure defined in the cipher text CT, the user can get the content key

V. IMPLEMENTATION

We implement a prototype of the authorized reduplication system, in which we are using JAVA Swings to develop the front end tool and MySQL is used as the backend tool. This tool has separate models such as

a) User program- user program is used to model the data user to carry out the file storing process.

b) Private server- private server program is mainly used manage the file token operation

c) Public server-public server program is used to model the public cloud which used to stores the large numbers of files.

d) Storage server-storage server progarm is used to model the storage-clod service provider which stores reduplicates the files.

Our implementation various tables are used to show the detail description of all users performed various operations in the cloud.

In the below flow chart, Owner is the one who is responsible for creating users under the cloud. The users added by the owner for the particular cloud server will be the authorized user to access the files from the cloud server. The Owner is solely responsible for uploading the files to the cloud server, User has read only permissions to that file by having the secret key provided by the owner. While uploading the file to the cloud server the file is checked in the cloud server for duplication, if file already stored the duplicate file will not be stored in the cloud. The user who is accessing the file from the cloud provides the secret key, owner name, filename to download the decrypted file to his local. The file stored in the cloud is encrypted using the AES algorithm. If authorized user enters the wrong secret key and try to access the file, the user will be blocked. Only owner can unblock the user.

In this paper cloud user that is client provides following function calls to support logo generation and reduplication along the file storing process

1) File-Logo(file) - It computes AES of the file as file-logo at the time of uploading the file.

ISSN: 2278 – 7798

2) DcheckReq(Logo) – It request the local public server to check duplicate copy of the file by sending file-logo received from the local server.
3) File-Encrypt(File) – It encrypts the file with confluent Encryption using AES algorithm .
4) FileStoreReq(File-ID, f-name, logo) - It uploads the file to the public server suppose if this file is unique at the time updates the stored file Logo.

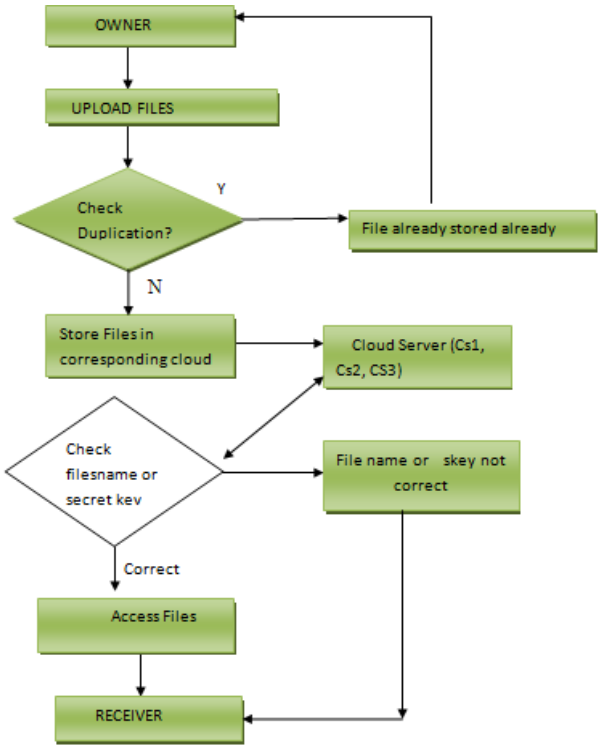

Fig.2. Flow Chart

In this implementation the local server includes related request for the logo generation and handles key storage with hash map. In public sever it provides re-duplication and file storage .It handles a map between existing file and associated logo with hash map.

VI. Demo Application

In this paper, finally in the result section we are fully eliminating the duplicate copies of the files and storing only a single copy of the original file, and also providing the security for that original file using confluent encryption method. Only authorized user can easily access the file from the storage server with the help of the key is stored in the data owner table. Finally in the result we are saving cost, time and bandwidth usage.

We are showing the result in the form of screenshots, as shown in the above fig 2 it shows the detail description of the result. Cloudserver1, cloud server2, cloud server3, connector, Receiver, and data-owner are the main domains in this paper and we are showing the result in a specific manner with the help of the screenshots for understanding purpose.

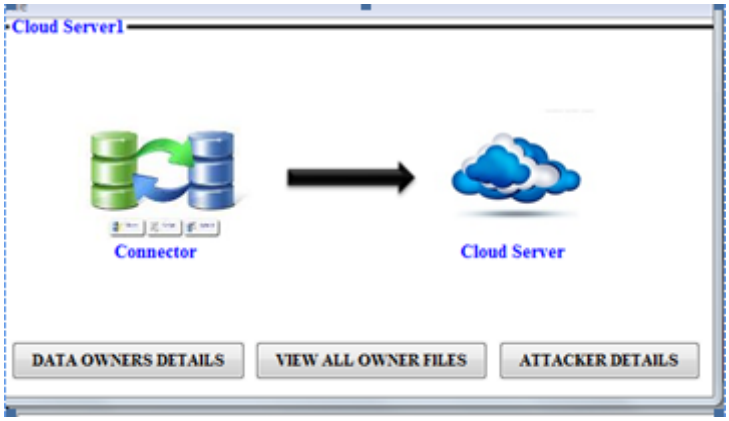

Fig3. Cloud Server

The above figure shows the cloud server, the connector in the figure is the one which is used to connect to the cloud. Any number of cloud servers can be used, and for every cloud the user have the separate access. Likewise any number of cloudServer can be created. The DataOwnerDetails button shows the onerdetails like name, organization etc. The viewAllOwnerFiles button shows the details of the files that are stored in the cloud. AttackerDetails stores the blocked user list. When the authorized user enters the wrong secret key, the user will be blocked .

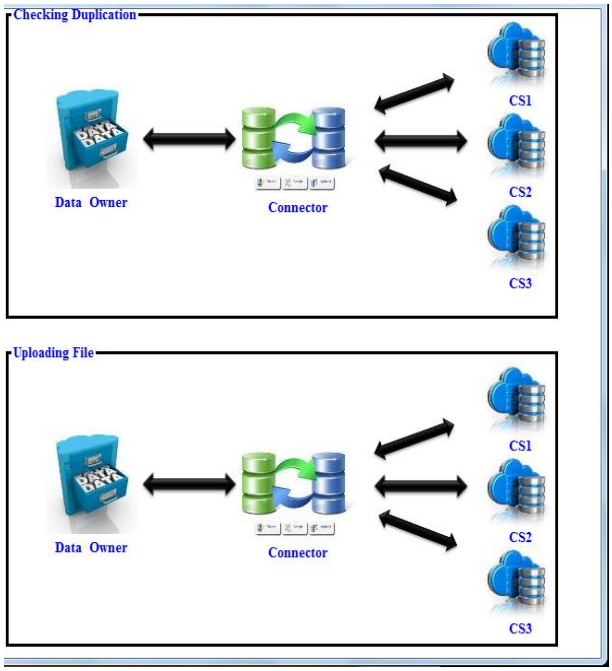

Fig. 4 connector

When the data owner tries to upload the file, it is checked for duplication. The owner selects the cloud server to which the file has to be uploaded. The file is checked for duplication in the particular cloud to which the user has to upload.

ISSN: 2278 – 7798

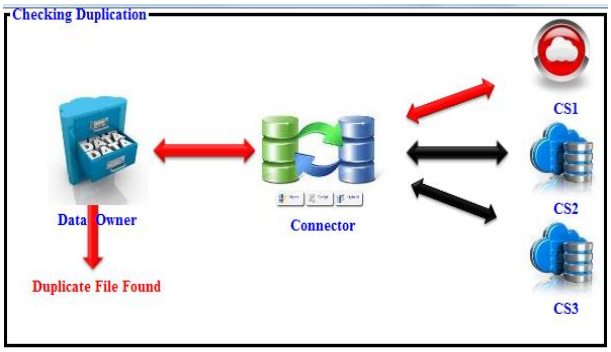

Fig. 5 duplicate file found

Once the file is found in the cloud server the server returns saying that the file is already present, so the file is not stored in the server. However if the user still has to upload the file he can clear the existing file and upload or he can upload it to the another cloud.

VII. CONCLUSION

In this paper, we showed how exactly to eliminate the duplicate copies of the files, and also to provide security using confluent encryption method. A user (data owner) derives a confluent key from each original data copy and encrypts the data copy with confluent key. In addition, the user also derives a logo for data copy, such that logo will be used to detect duplicates . As proof of idea, we showed that our secure duplex check method incurs to reduce the bandwidth, storage capacity, and also reduces amount of time. The confluent encryption method is less overhead compare to the normal encryption.

VIII. REFERENCES

[1] Q. He, Z. Li, and X. Zhang, "Data deduplication techniques," in *InternationalConference on Future Information Technology and ManagementEngineering (FITME)*, Changzhou, China, October 2010, pp. 430–433.

[2]. L. Aronovich, R. Asher, E. Bachmat, H. Bitner, M. Hirsch, andS. T. Klein, "The design of a similarity based deduplication system,"in*Proceedings of SYSTOR 2009: The Israeli Experimental SystemsConference*. New York, NY, USA: ACM, 2009, pp. 6:1–6:14.

[3]. L. DuBois, "Data deduplication for backup: Accelerating efficiency and driving down it costs," White paper, EMC Corporation, May 2009.

[4]. NetApp- Tech OnTap," Jun. 2009. [Online].Available: http://www.netapp.com/us/communities/tech-ontap/tot-dedupe-unstructure-0409.html

[5]. E. Bertino and R. Sandhu, "Database Security-Concepts,Approaches, and Challenges," IEEE Trans. Dependable and SecureComputing, vol. 2, no. 1, pp. 2-19, Jan.-Mar. 2005.

[6].Introduction to Cloud Computing,www.priv.gc.ca

[7]. Deduplication and Compression Techniques in Cloud Design,Amrita Upadhyay, Pratibha R Balihalli, ShashibhushanIvaturi and ShrishaRao,International Institute of Information Technology Bangalore Electronics City, Bangalore 560100.
Email: {amrita.upadhyay, pratibha.br, shashibhushan.ivaturisasi} @iiitb.org and shrao@ieee.org

[8]. A Secure Data Deduplication Scheme for Cloud Storage,(Jan Stanek, Alessandro Sorniottiy, Elli Androulakiy, and Lukas Kencl).

[9]. Fast and Secure Laptop Backups with Encrypted De-duplication(Paul Anderson University of Edinburghdcspaul@ed.ac.uk, Le Zhang niversity of Edinburghzhang.le@ed.ac.uk).

[10]. A Secure Cloud Backup System with Assured Deletion and Version Control.

[11].Security Proofs for Identity-Based Identification and Signature Schemes

[12].Twin Clouds: An Architecture for Secure Cloud Computing.

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

ISSN: 2278 – 7798