Secure Authorization of Data De-duplication using Hybrid Cloud Approach

Sayali K. Kamble[1] Rahul R. Patil[2] Shraddha S. Ashtekar[3]
[1,2,3]MIT Academy of Engineering

Abstract— Data de-duplication is one of major data compression techniques for removing duplicate copies of iterating data, and has been commonly used in cloud storage to diminish the amount of storage space and conserve bandwidth. In most organizations, the storage systems include duplicate copies of numerous pieces of data. For example, the identical file may be stored in several distinct places by distinct users, or two or more files that aren't identical may still contain much of the identical data. De-duplication removes these additional copies by saving only one copy of the data and substituting the other copies with pointers that escort back to the original copy. To preserve the confidentiality of delicate data while preserving de-duplication, the convergent encryption technique has been advanced to encrypt the data prior to outsource it. Different from existing de-duplication systems, the differential privileges of users can be contemplated in duplicate check except the data itself stated.

Key words: Cloud Computing, De-Duplication, Hybrid Cloud, Convergent

I. INTRODUCTION

A user is an entity that inclines to access the files from S-SCP. User creates the key and save that key in private cloud. In storage system under de-duplication. The user only uploads distinct data but do not upload any duplicate data to set aside the upload bandwidth, which may be possess by the user. Each file is safeguard by convergent encryption key and can access by only authorized person. In our system user must require to roster in private cloud for storing token with particular file which are store on public cloud. When user tends to access that file user access respective token from private cloud. After that user can access his files from public cloud. Token include file content D and convergent key KD.

A. Hybrid Cloud

1) Hybrid Cloud Approach for Security of User's Data.

Data deduplication problem is solved with differential privileges in cloud computing, hybrid cloud architecture [1] [7] is also being considered which is consisting of public cloud and private cloud. The system is also enhanced in the security. Thus the sensitive data of the users is being protected

Public cloud entity is utilized for the storage purpose. Public cloud is indistinguishable as S-CSP. When the user ought to download the files from public cloud, it will be query the key which is save in private cloud. When the users key is match with files key at that time user can download the file, unaccompanied by key user cannot ingress the file. Only permit user can ingress the file. In public cloud all files are ingress in encrypted format. If any possibility of unauthorized person hack our file, but without secrete or convergent key he doesn't ingress original file. On public cloud there are plenty of files are ingress each user access its respective file if it's token matches with S-CSP server token.

For data de-duplication hybrid cloud approach is used in which identical copies of stored data is eliminated. Main purpose is to utilize the storage space and also bandwidth of the network is saved. The management of the regularly enlarging amount of data is the critical provocation of cloud computing. Data de-duplication or isolated Instancing essentially mention to the deletion of inessential data. In the de-duplication action, only isolated copy (single instance) of the data is stored duplicate data is deleted. Nonetheless, indexing of all data is still kept such that data ever be required. In general the duplicate copies of repeating data is eliminated in data de-duplication. In computing, data de-duplication is a particular data compression expertise for eliminating duplicate copies of restating data. In the de-duplication process, distinctive lumps of data, or byte patterns, are recognized and cached during an operation of survey. Established encryption, while giving data confidentiality, is conflicting with data de-duplication. Different users encrypt their data with their own keys in traditional encryption method. Thus, de-duplication impossible as uniform data copies of different users will lead to non-identical cipher text .To impose data confidentiality while creating de-duplication feasible, Convergent encryption has been proposed .It encrypts/decrypts a copy of data by using a convergent key, which is derived under the aegis of evaluating the cryptographic hash value of the content of the data copy. However, previous de-duplication systems cannot reinforce *distinctive approval identical scan.* In this type of authorized de-duplication system, each user is provide a set of exemption during system initialization. Each file uploaded to the cloud is also bounded by a set of exemption to specify which kind of users is accessed to accomplish the duplicate check as well as access the files.

B. Data Duplication Problem

Storage efficiency functions such as de-duplication afford storage providers better utilization of their storage back ends and the ability to serve more customers with the same infrastructure. It is the process by which a storage provider only stores a single copy of a file owned by several of its users and there are four different de-duplication strategies, depending on whether de-duplication happens at the client side (i.e. before the upload) or at the server side, and whether de-duplication happens at a file level or at a block level. De-duplication is most rewarding when it is triggered at the client side, as it also saves upload bandwidth but For these reasons, de-duplication is a critical enabler for a number of popular and successful storage services which offers a cheap, remote storage to the broad public by performing client-side de-duplication, thus it will saving both the network bandwidth and storage costs. Indeed, data de-duplication is arguably one of the main reasons why the prices for cloud storage and cloud backup services have dropped so sharply.

II. ORGANIZATION

The paper prosecutes as follows. Section 3, it fleetingly revisit some preliminaries of this paper. In section 4, the architecture of de-duplication system is proposed. In Section

5, practical methods are proposed for de-duplication system with differential exemptions in cloud computing. The efficiency and security analysis for the suggested system are respectively dispense in Section 6. In section 7, the literature survey details are briefly explained further. Finally conclusion will drawn in Section 8

III. PRELIMINARIES

In this section, the functions are first defined and methods used, review some secure primitives used in our secure de-duplication.

A. Symmetric Encryption

Symmetric encryption operates on a secret key κ which is common to encrypt and decrypt information. There are three primitive functions for symmetric encryption.:

Key Gen SE(1_) ! κ Using security parameter secret key k generated using this algorithm 1_;

Enc SE(κ,M) ! C The symmetric encryption algorithm in which, the secret κ and message M are taken as inputs and then outputs the cipher text C.

Dec SE(κ,C) ! M is the symmetric decryption algorithm in which the secret κ and cipher text C are taken as input and then outputs the original message

B. Convergent Encryption

In de-duplication Convergent encryption provides data confidentiality. A convergent key is obtained by user (or data owner) from each original data copy and the data copy is encrypted accompanied by the convergent key. In addition, the user generate a label for the data copy, to detect replicates such that the tag will be used. Here, we assume that the tag correctness. It has been acquired for future property [4] holds, i.e., the labels are same for two same data copies. The detection of duplicates is done by the user by sending the tag to the server side to scan if the similar copy has been already stored. Note the derivation of both the convergent key and the label are independent, and to conclude the convergent key the tag cannot be used and conclude data confidentiality. On the server side the encrypted data copy as well as its respecive label will be stored. Formally, a convergent encryption technique can be explain with four primitive functions:

Key Gen CE(D) ! K: The key generation algorithm that assign a data copy D to a convergent key K;

Enc CE (K,D) ! C The symmetric encryption algorithm that takes both the convergent key K and the D as data copy inputs and then outputs a cipher text C;

Dec CE (K,C) ! D The decryption algorithm that takes input T as a cipher text and K as the convergent key and generates outputs the original data copy D;

Tag Gen (D) ! T(D) is the tag generation algorithm that plots the original data copy D and outputs a tag T(D).

IV. ARCHITECTURE

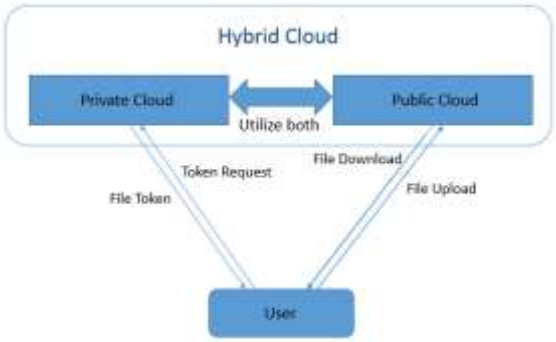

Fig. 1: Deduplication using hybrid cloud

V. PROJECT IDEA

The main idea is to provide the de-duplication of Files since the hybrid cloud which is a synthesis of the private cloud as well as public cloud. In common by if we used the public cloud we can't supply the security to our private data and thus our private data will be misplaced. Due to this we have to provide more security to our data. To achieve this we use a private cloud along with the public cloud. When we use private clouds the greater security can be provided which is used to refrain the replicate copies of data. User can transmit the files from public cloud but private cloud gives the security to the data. That incline only the authorized person can load and download the files from the public cloud for that user creates the key as well as save that key on the private cloud. At the identical time of downloading user query to the private cloud for key. After the response the user can access that certain file.

A. Authorization

Based on authority, individual token of files can accessed by each user for performing duplicate check. Under this assumption, a token for duplicate check cannot be generated by user without of his access or without the support from the private cloud server.

B. Authorized Duplicate Check:

Authorized user is capable to access own token which is stored on a private cloud, meanwhile the duplicate check is performed by the public cloud and the user is informed whenever there is any identical. The security demands considered is in two parts, the security of file token as well as to the data files. For the security of file token, two facets are defined as unforgeability and in distinguishability of file token. The details are given below.

C. Unforgeability of File Token

A registration is made by in private cloud for creating file token .Using particular file token, on public cloud user can upload or download files. The users are not permitted to collaborate accompanied by the public cloud server to crack the unforgeability of file tokens. In this system, the S-CSP is honest but inquisitive and will ethically perform the duplicate check upon receiving the identical invocation. The identical check token of users should be accoutre from the private cloud server in this strategy.

D. Indistinguishability of File Token

It requires that any user without questioning the private cloud server for each file token, he cannot get any helpful

information from the token, which embrace the file details as well as key details.

VI. SECURITY ANALYSIS

A. Security of Duplicate-Check Token:

Several types of protection needed is,

1) Unforgeability of Duplicate-Check Token:

As per the surveys there are two types of contenders, that is, external contender and internal contender. The external contender can be considered as an internal contender without any privilege. If a user has exemption p, it requires that the contender cannot forge and output a valid replicate token with any other privilege p′ on any file F, where p does not match p′. Moreover, it also needs that if the contender does not produce a desire of token with its own exemption from private cloud server, it cannot cast and out-turn a logical replicate token with p on any F that has been disputed. The internal contenders have more attack power than the external contenders and accordingly we only requre to consider the security averse to the internal attacker,

2) Incomprehensibility of Equivalent Check Token:

This characteristic is also stated in form of two types as the denotation of unforgeability. Initially, if a user has exemption p, given a token ϕ′, it requires that the contender cannot distinguish which exemption or file in the token if p does not match p′ .Moreover, it also needs that if the contender does not produce a desire of token with its own exemption from private cloud server, it cannot distinguish a valid replicate token accompanied by p on any other F that the contender has not disputed. In the security denotation of indistinguishablity, we need that the adversary is not permitted to intrigue with the public cloud servers. Actually, such a supposition could be discarded if the private cloud server conserves the list of the tags for all the files stored. There is an identical condition for the analysis of unforgeability, the security averse external contender is entail in the security averse the internal contenders.

VII. LITERATURE SURVEY

In [1], a technique for avoidance of replication of data on the cloud reposit has been suggested by the author. To protect the intimate of delicate data while supporting de-duplication, the convergent encryption technique is proposed

In [2], author proposed a system which achieves the data de-duplication by providing the evidence of data by the data owner. This evidence is used at the time of uploading of the file. A set of exemption bounds the file that is to be uploaded. Authorized users are permitted to execute the duplicate check and ingress the files.

From [3], in this paper author has put forward the leading ways to removal of all those problems by taking into consideration hybrid cloud architecture, in which public cloud makes available to data owner for providing storage place which will managed by private cloud act as an proxy to allow data owner and user with security and privacy along with different privileges set.

In [4], Author has proposed distinctive authentication duplicate check which is most significant in application in that authorized duplicate system .A set of exemptions is given to each user in such system.

In [5], this proof of ownership protocol implies for halting unapproved access to provide proof that the user certainly owns the similar file when de-duplication found. For accessing the file at time of file uploading a pointer is provided to the user. Similarly for the download operation the user downloads the encrypted file and for decryption the convergent key is used.

In [6], the problem of achieving efficient as well as reliable key management in secure de-duplication is addressed by author from the respective paper. A standard approach is introduced by the author in which an autonomous master key is held by each user for encrypting the convergent keys and outsourcing them.

VIII. CONCLUSION

In the traditional system there was an issue of data duplication. Redundant copies of data are eliminated provided that, the data security is enhanced due to technique of de-duplication. Along with the security the network transfer rate is reduced.

REFERENCES

[1] Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou, “A Hybrid Cloud Approach for Secure Authorize De-duplication”. Volume: PP, Issue:99, Date of Publication :18.April.2014.

[2] Gaurav Kakariya, Prof. Sonali Rangdale,” A hybrid cloud approach for secure Authorized de-duplication”. Volume VIII, Issue I, October 14

[3] Sharma Bharat, Mandre B.R.,”A Secured and Authorized Data De-duplication in Hybrid Cloud with Public Auditing”, International Journal of Computer Applications (0975 – 8887) Volume 120 – No.16, June 2015.

[4] Mr.Vinod B Jadhav, Prof.Vinod S Wadne, “ Secured Authorized De-duplication Based Hybrid Cloud Approach”, Volume 4, Issue 12, December 2014

[5] Rajashree Shivshankar Walunj, Deepali Anil Lande, Nilam Shrikrushna Pansare,”Secured Authorized De-duplication Based Hybrid cloud“. The International Journal of Engineering and Science (IJES), Volume- 3, Issue-11, 2014.

[6] Nikhil O. Agrawal, Prof.S.S.Kulkarni, ”Secure Deduplication and Data Security with Efficient and Reliable Convergent Key Management”, International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 7, July 2015

[7] Prajakta Patil, Mr. Anilkumar Warad, “A Survey on Data De duplication Techniques”, Int.J.Computer Technology & Applications,Vol 5 (6),1964-1967.

