Integrity proofes for de-duplication on hybrid cloud computing

Prof.D.G.Sancheti
Lecturer
Computer Engineering,
S.N.J.B's KBJ COE, Chandwad, sancheti.dgcoe@snjb.org

Abstract:
Data De-duplication can be used for eliminating duplicate data on cloud and is one of key data compression method. This helps to reduce the data storage space and save network bandwidth. To encrypt the data convergent encryption technique has been developed, and it also helps to preserve confidentiality of data .The differential privileges of users are further considered in duplicate check besides the data itself is different from traditional De-duplication systems. Authorized duplicate check in hybrid cloud architecture is constructed by the system which provides several new De-duplications. The authorized duplicate check scheme is proposed by a prototype which implement by system.

IndexTerms **-** De-duplication, Authorized Duplicate Check, Confidentiality, Hybrid Cloud.

I .INTRODUCTION

The term "Cloud" originates from the World of telecommunications when providers began using virtual private network (VPN) services for data communications. The definition of cloud computing provided by National Institute of Standards and Technology (NIST). Cloud computing is an on demand service in which shared resources, information, software and other devices are provided according to the clients requirement at specific time. It's a term which is generally used in case of Internet. The whole Internet can be viewed as Cloud. Capital and operational costs can be cut using Cloud computing. In a different files at the block level the data deduplication can also be occur, so elimination of duplicate block can be done. By relevant data owners with their convergent keys can only able to decrypt the file which where downloaded by the user which is in encrypted form with the help of pointer from the server. In this unauthorized user who tries to access the file is prevented. Based on the domain or environment, in which clouds are used, can be divided into 3 types those are: Public Cloud, Private Cloud, and Hybrid Cloud.

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, De-duplication has been a well-known technique and has attracted more and more attention recently. Data De-duplication is a specialized data compression technique for eliminating duplicate copies of repeating data in storage.

II. LITERATURE SURVEY

Neal Leavitt, 2013, has mentioned that: Hybrid Cloud is the architecture that provides the organization to efficiently work on both the private and public cloud architecture in combination by providing the scalability to adopt. Here some of the basic concepts and idea proposed by authors and how best and easy to adopt this environment [1].

Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, WenjingLou, 2014, have mentioned that: System is designed to solve the deferential privilege problem in secure Deduplication. The security will be analyzed in terms of two aspects, that is, the authorization of duplicate check and the confidentiality of data. Some basic tools have been used to construct the secure Deduplication, which are assumed to be secure. In these Systems calculate the signature value to verification of redundant data, using these System can be easily identify the duplicate data on the cloud. These basic tools include the convergent encryption scheme, symmetric encryption scheme [3].

V SreekarBabu, 2014, has mentioned that: De-duplication is a technique used for removing repeating copies of data and used in cloud storehouse to reduce the amount of storage space. In order to protect the private data of users while supporting De-duplication, Convergent encryption has been used to encrypt the data. To better protect data security, in this proposed work, a set of privileges are issue to each user and each file uploads to the cloud and the deferential privileges of both user and files are further consider in duplicate check. In the proposed work, the duplicate blocks of data in non-identical files can also be detectable with the help of Meta data manager [4].

Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, WenjingLou, 2014, have mentioned that: Proposed system is designed to solve the differential privilege problem in secure Deduplication. The security will be analyzed in terms of two aspects, that is, the authorization duplicate check and the confidentiality of data. Some basic tools have been used to construct the secure Deduplication, which are assumed to be secure. In these Systems calculate the signature value to verification of redundant data; using these Systems

can be easily identify the duplicate data on the cloud. These basic tools include the convergent encryption scheme, symmetric encryption scheme [5].

In the Data De-duplication product review, system knows about the existing system. In which the total overall survey of existing system, which hardware and software are used, which algorithms are used, which processes are used in the previous system.So,from these review system presents new ideas in these project.

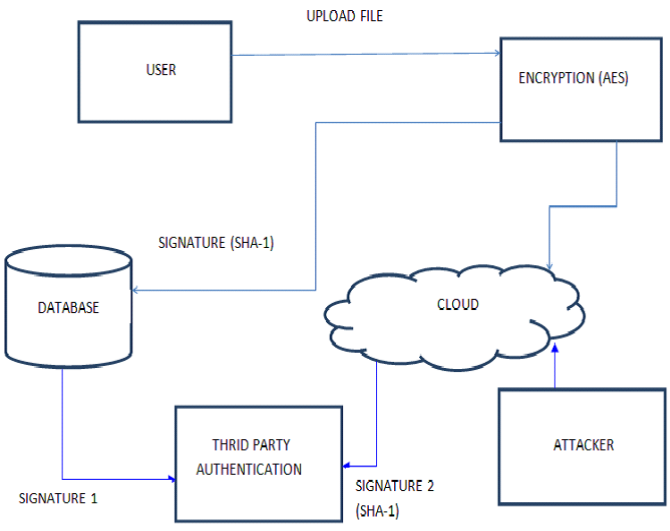
V SreekarBabu, 2015, has mentioned that: De-duplication is a technique used for removingre1peating copies of data and used in cloud storehouse to reduce the amount of storage space. In order to protect the private data of users while supporting De-duplication, Convergent encryption has been used to encrypt the data. To better protect data security, in this proposed work, a set of privileges are issue to each user and each file uploads to the cloud and the differential privileges of both user and files are further consider in duplicate check. In the proposed work, the duplicate blocks of data in non-identical files can also be detectable with the help of Meta data manager [2].

III. PROPOSED SYSTEM

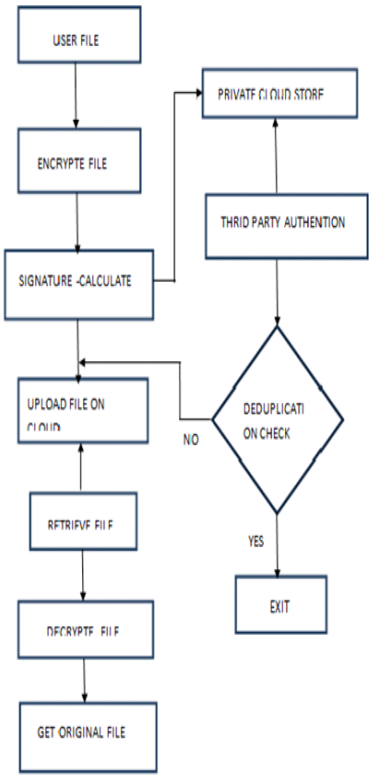
A proposed system defines as de-duplication can take place at either the file level or the block level. For file level De-duplication, it eliminates duplicate copies of the same file. De-duplication can also take place at the block level, which eliminates duplicate blocks of data that occur in non-identicalness'. Although data De-duplication brings a lot of bents, security and privacy concerns arise as users' sensitive data are susceptible to both inside and outside attacks. Traditional encryption, while providing data confidentiality, is incompatible with data De-duplication.

Specifically, traditional encryption requires different users to encrypt their data with their own keys. Thus, identical data copies of different users will lead to different cipher texts, making De-duplication impossible. Convergent encryption has been proposed to enforce data confidentiality while making De-duplication feasible. It encrypts/decrypts a data copy with a convergent key, which is obtained by computing the cryptographic hash value of the content of the data copy. After key generation and data encryption, users retain the keys and send the cipher text to the cloud. Since the encryption operation is deterministic and is derived from the data content, identical data copies will generate the same convergent key and hence the same cipher text. To prevent unauthorized access, a secure proof of ownership protocol is also needed to provide the proof that the user indeed owns the same file when a duplicate is found. After the proof, subsequent users with the same le will be provided a pointer from the server without needing to upload the same le.

A user can download the encrypted file with the pointer from the server, which can only be decrypted by the corresponding data owners with their convergent keys. Thus, convergent encryption allows the cloud to perform De-duplication on the cipher texts and the proof of ownership prevents the unauthorized user to access the file. However, previous De-duplication systems cannot support differential authorization duplicate check, which is important in many applications. In such an authorized De-duplication 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 toper form 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. 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.



System Architecture



System Design Data Flow Diagram

REFERENCE

1. NealLeavitt, "Hybrid Clouds Move to the Forefront" Published by the IEEE Computer Society, MAY 2013

2. V SreekarBabu," An Authorized Duplicate Check in a Hybrid Cloud Architecture", September 2015.

3.Jin Li, Yan Kit Li, Xiao Feng Chen, Patrick P. C. Lee, WenjingLou, "A Hybrid Cloud Approach for Secure Authorized De-duplication", IEEE Transactions On Parallel and Distributed System vol:pps no:99 year 2014.

4. V SreekarBabu," An Authorized Duplicate Check in a Hybrid Cloud Architecture", September 2015

5.Jin Li, Yan Kit Li, Xiao Feng Chen, Patrick P. C. Lee, WenjingLou, "A Hybrid Cloud Approach for Secure Authorized De-duplication", IEEE Transactions On Parallel and Distributed System vol:pp no:99 year 2014.

