SAVING THE STORAGE SPACE IN CLOUD COMPUTING BY USING DEDUPLICATION MECHANISM

NIHARIKA MANDA
M.S (computer information systems and information technology)
University of central Missouri

ABSTRACT:

In individualized computing devices that rely upon a conveyed stockpiling condition for data support, a quick moving toward test going up against source de-duplication for cloud fortification organizations is the low de-duplication viability due to a mix of the benefit heightened nature and the obliged structure resources. Data de-duplication is one of basic data weight methodologies for discarding duplicate copies of repeating data, and has been for the most part used as a piece of appropriated stockpiling to reduce the measure of capacity room and extra transmission limit. To secure the mystery of fragile data while supporting de-duplication, the combined encryption technique has been proposed to encode the data previously outsourcing. To better guarantee data security, this paper makes the principle try to formally address the issue of endorsed data de-duplication. Not exactly the same as customary de-duplication structures, the differential advantages of customers are additionally considered in duplicate check other than the data itself. We furthermore demonstrate a couple of new de-duplication improvements supporting affirmed duplicate check in cross breed cloud outline. Security examination demonstrates that our arrangement is secure the extent that the definitions decided in the proposed security display. As a proof of thought, we execute a model of our proposed affirmed duplicate check design and direct attempted investigations using our model. We exhibit that our proposed endorsed duplicate check design obtains irrelevant overhead appeared differently in relation to ordinary activities.

Keywords: *de-duplication, cloud computing*

1. INTRODUCTION:

Appropriated figuring gives clearly limitless "virtualized" advantages for customers as organizations over the whole Web, while disguising stage and execution unobtrusive components. The present cloud organization providers offer both significantly open limit and massively parallel enlisting resources at by and large low costs. As circulated processing gets the chance to be normal, an extending measure of data is being secured in the cloud and granted by customers to decided advantages, which describe the passageway benefits of the set away data. One essential trial of dispersed capacity organizations is the organization of the ceaselessly extending volume of data. To make data organization flexible in disseminated figuring, de-duplication has been a without a doubt comprehended technique furthermore, has pulled in more thought starting late. Data de-duplication is a particular data weight technique for discarding duplicate copies of repeating data away. The technique is used to upgrade amassing use furthermore, can in like manner be associated with organize data trades to diminish the amount of bytes that must be sent.

As opposed to keeping various data copies with a similar substance, de-duplication takes out overabundance data by keeping one and just physical copy and insinuating other monotonous data to that copy. De-duplication can occur at either the report level or the piece level. For archive level de-duplication, it takes out duplicate copies of a similar record. De-duplication can in like manner occur at the piece level,

which takes out duplicate squares of data that occur in non-unclear records. Conveyed figuring is a creating organization shows that gives computation and limit resources on the Internet. One engaging value that dispersed registering can offer is disseminated capacity. Individuals and endeavors are routinely required to remotely record their data to remain from any information disaster if there are any gear/programming disillusionments or unexpected calamities. Instead of purchasing the required accumulating media to keep data fortifications, individuals what's more, endeavors can essentially outsource their data support organizations to the cloud organization providers, which give the fundamental accumulating advantages for have the data fortifications. While circulated capacity is engaging, how to give security affirmations to outsourced data transforms into a rising concern. One vital security test is to give the property of ensured eradication, i.e., data records are forever blocked perpetual supply of cancelation. Keeping data fortifications forever is bothersome, as delicate information may be revealed later because of data burst or mixed up organization of cloud overseers. Along these lines, to keep up a vital separation from liabilities, endeavors and government associations as a general rule keep their fortifications for a set number of years and sales to delete (or obliterate) the fortifications from that point. For occasion, the US Congress is itemizing the Internet Data Retention authorization in moving toward ISPs to hold data for quite a while, while in United Kingdom, associations are required to hold wages and pay records for a long time.

2. OBJECTIVES AND GOALS:

1. Unforged ability of file token/duplicate-check token. Unauthorized users without appropriate privileges or file should be prevented from getting or generating the file tokens for duplicate check of any file stored at the SCSP. The users are not allowed to collude with the public cloud server to break the unforged ability of file tokens. In our system, the S-CSP is honest but curious and will honestly perform the duplicate check upon receiving the duplicate request from users. The duplicate check token of users should be issued from the private cloud server in our scheme.
2. In distinguishbility of file token/duplicate-check token. It requires that any user without querying the private cloud server for some file token, he cannot get any useful information from the token, which includes the file information or the privilege information.
3. Data Confidentiality. Unauthorized users without appropriate privileges or files, including the S-CSP and the private cloud server, should be prevented from access to the underlying plaintext stored at S-CSP. In another word, the goal of the adversary is to retrieve and recover the files that do not belong to them. In our system, compared to the previous definition of data confidentiality based on convergent encryption, a higher level

confidentiality is defined and achieved.

OBJECTIVE:

1. To improved integrity
2. To increase the storage utilization
3. To remove the duplicate copies of data and improve the reliability.
4. To improve the security

3. LITERATURE SURVEY:

Data de-duplication is a technique for reducing the amount of storage space an organization needs to save its data. In most organizations, the storage systems contain duplicate copies of many pieces of data. For example, the same file may be saved in several different places by different users, or two or more files that aren't identical may still include much of the same data. De-duplication eliminates these extra copies by saving just one copy of the data and replacing the other copies with pointers that lead back to the original copy. Companies frequently use de-duplication in backup and disaster recovery applications, but it can be used to free up space in primary storage as well. To avoid this duplication of data and to maintain the confidentiality in the cloud we using the concept of Hybrid cloud. To protect the confidentiality of sensitive data while supporting de-duplication, the convergent encryption technique has been proposed to encrypt the data before outsourcing. To better protect data security, this paper makes the first attempt to formally address the problem of authorized data de-duplication. [5]

Data de-duplication is one of important data compression techniques for eliminating duplicate copies of repeating data, and has been widely used in cloud storage to reduce the amount of storage space and save bandwidth. To protect the confidentiality of sensitive data while supporting de-duplication, the convergent encryption technique has been proposed to encrypt the data before outsourcing. To better protect data security, this paper makes the first attempt to formally address the problem of authorized data de-duplication. Different from traditional de-duplication systems, the differential privileges of users are further considered in duplicate check besides the data itself. We also present several new de-duplication constructions supporting authorized duplicate check in a hybrid cloud architecture. Security analysis demonstrates that our scheme is secure in terms of the definitions specified in the proposed security model. As a proof of concept, we implement a prototype of our proposed authorized duplicate check scheme and conduct test bed experiments using our prototype. We show that our proposed authorized duplicate check scheme incurs minimal overhead compared to normal operations.[3]

Data de-duplication is one of important data compression techniques which are for eliminating duplicate copies of repeating data, and has been widely used in cloud storage in order to reduce the amount of storage space and save bandwidth. To protect the confidentiality of sensitive data while supporting de-duplication, the convergent encryption technique has been proposed to encrypt the data before outsourcing. To better protect data security, papers makes the first attempt to formally address the problem of authorized data de-duplication. Different from traditional deduplication systems, the differential privileges of users are further considered in

duplicate check besides the data itself. Also present several new de-duplication constructions supporting authorized duplicate check in hybrid cloud architecture. Security analysis demonstrates that the proposed scheme is secure in terms of the definitions specified in the proposed security model. [2]

This paper represents that, many techniques are using for the elimination of duplicate copies of repeating data, from those techniques, one of the important data compression technique is data duplication. Many advantages with this data duplication, mainly it will reduce the amount of storage space and save the bandwidth when using in cloud storage. To protect confidentiality of the sensitive data while supporting de-duplication data is encrypted by the proposed convergent encryption technique before out sourcing. Problems authorized data duplication formally addressed by the first attempt of this paper for better protection of data security. This is different from the traditional duplication systems. The differential privileges of users are further considered in duplicate check besides the data itself. In hybrid cloud architecture authorized duplicate check supported by several new duplication constructions. Based on the definitions specified in the proposed security model, our scheme is secure. Proof of the concept implemented in this paper by conducting test-bed experiments. [6]

The popularity and widespread use of Cloud have brought great convenience for data sharing and data storage. The data sharing with a large number of participants take into account issuers like data integrity, efficiency and privacy of the owner for data. In cloud storage services one critical challenge is to manage ever increasing volume of data storage in cloud. To make data management more scalable in cloud computing field, de-duplication a well-known technique of data compression to eliminating duplicate copies of repeating data in storage over a cloud. Even if data de-duplication brings a lot of benefits in security and privacy concerns arise as user's sensitive data are susceptible to both attacks insider and outsider. A convergent encryption method enforces data confidentiality while making de-duplication feasible. Traditional de-duplication systems based on convergent encryption even though provide confidentiality but do not support the duplicate check on basis of differential privileges. This paper presents, the idea of authorized data deduplication proposed to protect data security by including differential privileges of users in the duplicateCheck. [7]

4. EXISTING APPROACH:

- Data deduplication systems are the systems which can acts as a proxy server and it can allow the data users or owners to perform the duplicate check with using different types of privileges and it is secure.
- 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.

DISADVANTAGES OF EXISTING SYSTEM:

- Traditional encryption, while providing data confidentiality, is incompatible with data deduplication.
- Identical data copies of different users

will lead to different ciphertexts, making deduplication impossible.

5. ALGORITHMS:

Convergent Encryption:

Convergent encryption provides data confidentiality in de-duplication. A user (or data owner) derives a convergent key from each original data copy and encrypts the data copy with the convergent key. In addition, the user also derives a tag for the data copy, such that the tag will be used to detect duplicates. Here, we assume that the tag correctness property holds, i.e., if two data copies are the same, then their tags are the same. To detect duplicates, the user first sends the tag to the server side to check if the identical copy has been already stored. Note that both the convergent key and the tag are independently derived and the tag cannot be used to deduce the convergent key and compromise data confidentiality. Both the encrypted data copy and its corresponding 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 generation algorithm that maps a data copy M to a convergent key K.
2. EncCE(K, M)!C is the symmetric encryption algorithm that takes both the convergent key K and the data copy M as inputs and then outputs a ciphertextC;
3. DecCE(K, C)!M is the decryption algorithm that takes both the ciphertextC and the convergent key K as inputs and then outputs the original data copy M; and
4. TagGen(M)!T (M) is the tag generation algorithm that maps the original data copy M and outputs a tag T (M).

2) Proof Of Ownership

The notion of proof of ownership(POW) enables users to prove their ownership of data copies to the storage server.Specifically, POW is implemented as an interactive algorithm (denoted by POW). The verifier derives a short value $\phi(M)$ from a data copy M. To prove the ownership of the data copy M, the properneeds to send $\phi$to the verifier such that $\phi = \phi(M)$.

PSEUDO CODE

Step1:Calculate the two convergent key values

Step2: Compare the two keys and files get accessed. Step3: Apply de-duplication to eradicate the duplicate values.

Step4: Ifany other than the duplicates it will be checked once again and make the data unique.

Step5: That data will be unique and also more confidential the authorized can access and data is stored.

6. CONCLUSION AND FUTURE WORK

A few new de-duplication developments supporting approved copy check in half and half cloud engineering, in which the copy check tokens of documents are produced by the private cloud server with private keys. Security examination exhibits that our plans are secure as far as insider and pariah assaults indicated in the proposed security model. As a proof of idea, we executed a model of our proposed approved copy check plan and direct proving ground investigates our model. We demonstrated that our approved copy check plan acquires insignificant overhead contrasted with united encryption and system exchange.

7. REFERENCES

1. S. Plank LihaoXu Optimizing Cauchy Reed-Solomon Codes for Fault-Tolerant Network Storage Applications The 5th IEEE International Symposium on Network Computing and Applications (IEEE NCA06), Cambridge, MA, July, 2006

2. Mr Vinod B Jadhav Prof Vinod S Wadne Secured Authorized De-duplication Based Hybrid Cloud Approach International Journal of Advanced Research in Computer Science and Software Engineering

3. Abdul Samadhu, J. Rambabu, R. Pradeep Kumar, R. Santhya Detailed Investigation on a Hybrid Cloud Approach for Secure Authorized Deduplication International Journal for Research in Applied Science and Engineering Technology (IJRASET)

4. JadapalliNandini, Rami reddyNavateja Reddy Implementation De-duplication System with Authorized Users International Research Journal of Engineering and Technology (IRJET)

5. Sharma Bharat, Mandre B.R. A Secured and Authorized Data De-duplication with Public Auditing International Journal of Computer Applications (09758887)

6. Wee Keong Ng SCE, NTU Yonggang Wen SCE, NTU Huafei Zhu Private Data De-duplication Protocols in Cloud Storage SAC12 March 2529, 2012, Riva del Garda, Italy. Copyright 2011 ACM 9781450308571/12/03

7. Shweta D. Pochhi, Prof. Pradnya V. Kasture Encrypted Data Storage with De-duplication Approach on Twin Cloud International Journal of Innovative Research in Computer and Communication Engineering

8. Backialakshmi. N Manikandan. M SECURED AUTHORIZED DE-DUPLICATION IN DISTRIBUTED SYSTEM IJIRST International Journal for Innovative Research in Science and Technology— Volume 1 — Issue 9 — February 2015

9. BhushanChoudhary, AmitDravid A Study On Secure Deduplication Techniques In Cloud Computing International Journal of Advanced Research in Computer Engineering and Technology (IJARCET) Volume 3, Issue 12, April 2014