www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 4 Issue 2 February 2015, Page No. 10493-10497

A Novel Approach For Authorized Deduplication

Anup V. Adke[1] Akash N. Nawale[2] Pravin N. Jadhav[3] Vijayandra A. Yeole[4] Prof. Kavita S. Kumavat[5]

[1]Student of BE Information Technology
BVCOE & RI, Nasik, Maharashtra, India
University of Pune
anupadke8393@gmail.com

[2]Student of BE Information Technology
BVCOE & RI, Nasik, Maharashtra, India
University of Pune
akashnawale888@gmail.com

[3]Student of BE Information Technology
BVCOE & RI, Nasik, Maharashtra, India
University of Pune
pravinjadhav.pnj@gmail.com

[4]Student of BE Information Technology
BVCOE & RI, Nasik, Maharashtra, India
University of Pune
vijayandra.yeole@gmail.com

[5]ME Computer Engineering
BVCOE & RI, Nasik, Maharashtra, India
University of Pune
kavitakumavat26@gmail.com

Abstract: ***Now a day's data deduplication is one of the relevant data compression techniques for removing duplicate copies of repeating data, and has been broadly used in cloud storage to reduce the amount of storage space and save bandwidth. To defend the privacy of sensitive data while supporting unrepeatable, the convergent encryption procedure has been suggested to encrypt the data before outsourcing. To well protect data confidence, first attempt to officially address the problem of authorized data deduplication. Dissimilar from out-of-date deduplication organizations, the variance rights of users are further considered in duplicate check besides the data itself. We also present some new deduplication constructions supporting authorized duplicate check in hybrid cloud architecture. Security analysis proves that our scheme is secure in terms of the definitions indicated in the proposed security model. As evidence, implementation contains a prototype of our proposed authorized duplicate check scheme and conduct tested experiments using our prototype. System show that our suggested legal replacement check scheme acquires minimal overhead compared to normal operations.***

Keywords –AES Algorithm, SHA-1 Algorithm, Encryption, Decryption, Security.

I. Introduction

A Hybrid Cloud approach terms that it is the cloud service providers which offer high storage and parallel computing resource at relatively low cost .cloud computing provide apparently unlimited "virtualized" resources to users as services across the complete Internet, while hiding stand and operation details. Today's cloud facility suppliers agreement both highly obtainable loading. As cloud calculating has been spread broadly, an increasing amount of data is being stored in the cloud and shared by users with specified rights, which describe the admission privileges of the kept data. Due to the increasing data volume the cloud storage services are faced judgmentally .to make the data organization calmer in the cloud computing, deduplication is been well - known technique and has gained more an more attention now a days. Data deduplication is the well technique for eliminating the duplicate copies of repeating data in storage. This technique is used to improve the storage utilization and can also be applied to network data transfer to reduce the number of bytes that are to be sent. Instead of keeping multiple data copies with the same content, deduplication eliminates out of a job data by keeping only one physical copy and referring other terminated data copy.

Even though data deduplication brings a portion of incomes, security and confidentiality worries arise as user's sensitive data are vulnerable to both insider and outsider outbreaks. Outdated encryption, while as long as data privacy, and is unsuited with data deduplication. Exactly, traditional encryption requires different users to encrypt their data with their own keys. Thus, equal data copies of different users will lead to dissimilar encryption texts, creation deduplication tough. Convergent encryption has been planned to enforce data confidentiality while making deduplication possible. It encrypts/ decrypts a data duplicate with a convergent key, which is acquired by computing the cryptographic hash value of the content of the data copy. After key group and data cipher, users maintain the keys and send the cipher text to the cloud. Since the cipher process is deterministic and is resulting from the data loading, same data copies will obtain the same convergent key and hence the same cipher text. To defend illegal admission, a protected proof of proprietorship rules is also needed to provide the proof that the user definitely owns the similar folder when a replacement is found. After the proof, following users with the same file will be provided a pointer from the server without needing to upload the same file. A user can transfer the encoded folder with the pole from the server, which can only be decrypted by the matching data owners with their convergent keys. Thus, merged cipher agrees the cloud to achieve deduplication on the cipher texts and the proof of ownership avoids the unauthorized user to access the file.

However, previous deduplication systems cannot support differential authorization duplicate check, which is important in many application. In such an authorized deduplication system, each used is allotted a set of privileges during organization beginning. All folder uploaded to the cloud is also bounded by set of privileges to specify which kind of user is acceptable to perform the duplicate check and right to use the files. Already defer to his replacement check invitation for some file , the user needs to take file an his own privileges an inputs .The user is able to find a duplicate for this file if an only if there is a copy if this file and a corresponding privilege stored in cloud . For example, in a company many different privileges will be assigned to employee.

In order to save cost an efficiently management , the data will be moved to the storage server provider in the public cloud with identified privileges the deduplication techniques will be applied to store only in copy of the same file . Because of privacy concern, some files will be encrypted an acceptable the duplicate check by employee with specified privileges to realize the access control. Traditional, deduplication system based on merging cipher, although given that privacy to certain amount, do not provision the replacement check with variance privileges in other word, no differential privileges have been considered in deduplication based on convergent encryption technique. Its seems to be reversed if we want to recognise both deduplication and differential authorization duplicate check at the same time.

II. LITERATURE SURVEY

Literature survey is the most important step in software development process before developing the stool it is necessary to determine the time factor, budget n corporation power. Once these things are fulfilled, then next steps are to control which operating system a verbal can be used for developed the stool [1]. Once the programmers start construction the stool program writer need lot of external provision. This support can be gained from senior program writer, from book or from website previously construction the system the above concerns are taken into account for developed the proposed system [2]. In this paper, aiming at efficiently solving the problem of deduplication with differential privileges in cloud calculating, we study a mixture cloud construction holding of a public cloud and a private cloud[3]. Dissimilar present data deduplication structures, the isolated cloud is involved as a proxy to allow data owner/users to strongly perform replacement checked with difference rights. Such architecture is useful and has attracted much respect from investigators[4].The data proprietors only subcontract their data storage by using public cloud while the data operation is succeeded in isolated cloud. A new deduplication system supporting differential duplicate check is proposed under this hybrid cloud architecture where the S-CSP exists in the free cloud. The user is only permissible to achieve the duplicate check for files noticeable with the corresponding privileges [5].

Furthermore, we enhance our structure in safety. Exactly, we current a progressive arrangement to support stronger security by encrypting the file with differential honor answers. In this method, the customers without parallel rights cannot achieve the duplicate check. Furthermore, such illegal users cannot decrypt the cipher text even collude through the S-CSP [6].

Table 1: Symbol Description

Acronym	Description
S-CSP	Storage-cloud service provider
PoW	Proof of Ownership
(pk_U,sk_U)	User’s public and secret key pair
Kf	Convergent encryption key for file *F*
PU	Privilege set of a user *U*
PF	Specified privilege set of a file *F*
$\phi'F,p$	Token of file *F* with privilege *p*

Symbolizations Used in this table establishes that our scheme is safe in terms of the meanings stated in the projected security model. In existing system data duplication is one of important data compression technique 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 confidentiality of sensitive data will supporting reduplication, cloud calculating afford seemingly unlimited “virtualized “resource to users as facilities across the hole internet, will hiding platform and execution specifics. Today’s cloud service suppliers proposal together highly available storage and massively parallel computing resounds at relevantly low costs. as cloud computing become prevalent, an increasing amount of data is being stored in the cloud an share by users with specified privileges, which define the access rights of stored data .
Problems of existing system:
The managing of the always growing size of data is the serious experiment to the cloud service provider.

Proposed System:

System consist the three phase
1) System setup
2) File uploading
3) File retrieving

In the first phase private cloud server maintains a table which stores each user's identity and its corresponding privilege. In the second phase the user uploads file on the cloud if the file is found to ne duplicate the cloud asks for the authority or identification to the user an if the file is not duplicate then the cloud sends the token to the user in the form of signature .an then the file is uploaded. In the third phase the procedure of file retrieving is similar to the construction.

III. System Overview

A novel architecture for data deduplication in haze computing, which consists of a twin clouds (i.e. the public cloud and the private cloud). Really, this mixture cloud situation has involved more and more attention newly. For example, an enterprise might use a open cloud capability, such as Amazon S3, for archived data, but carry on to save up in-house filling for working client data. Instead, the confidential isolated cloud be a cluster of virtualized cryptographic co-processors, which are obtainable as a facility by a third gathering and afford the necessary hardware based security features to implement a remote execution environment trusted by the users.

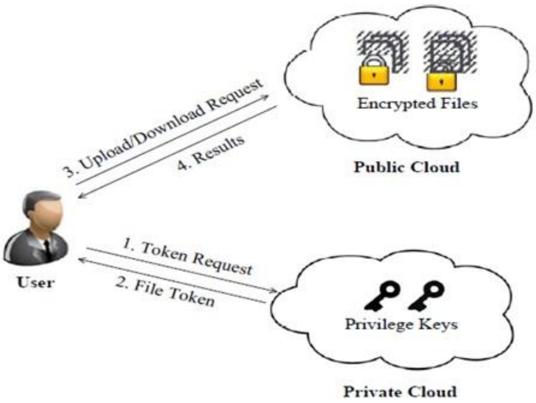

Figure 1:System Architecture

At a high level, our setting of attentiveness is an enterprise network, consisting of a group of united clients who will usage the S-CSP and stock data with deduplication method. In this location, deduplication can be often used in these locations for data backup and tragedy recovery applications while greatly falling storing space. Such systems are general and are frequently more appropriate to user file backup and synchronization applications than richer storage generalizations. There are three objects definite in our scheme, that is, users, isolated cloud and S-CSP in open cloud as presented in Figure 1. The S-CSP performs deduplication by examination if the subjects of two files are the similar and stores individual one of them.

- S-CSP: This is an entity that offers a data storage service in open cloud. The S-CSP offers the data farm out facility and stores data on behalf of the users. To decrease the storage cost, the S-CSP eliminates the storage of redundant data via deduplication and keeps one single data. In this paper, we take responsibility that S-CSP is continuously connected and has abundant storage volume and calculation power.

- Data Users: A user is an entity that wants to outsource data storage to the S-CSP and access the data later. In a storage system supportive deduplication, the user one uploads single data but does not upload any replacement data to save the upload bandwidth, which may be possessed by the same user or dissimilar users. In the official deduplication structure, each user is delivered a set of rights in the format of the system. Each file is secure with the merging cipher key and license keys to understand the official deduplication with variance rights.

- Private Cloud: Compared with the traditional deduplication architecture in cloud computing, this is a new entity introduced for simplifying user's secure usage of cloud facility. Exactly, since the calculating assets at data user/proprietor side are limited and the public cloud is not fully right-hand in exercise, isolated cloud is able to offer data user/proprietor with an implementation environment and substructure employed as a boundary between user and the open cloud. The isolated answers for the rights are achieved by the isolated cloud, who responses the file nominal demands from the users. The boundary obtainable by the isolated cloud permits user to submit files and queries to be securely stored and calculated correspondingly.

  By using this we are trying to overcome the critical challenge of cloud storage services for management of the ever-increasing volume of data.

IV. Algorithmic Strategy

- **Symmetric Encryption Algorithm:** Symmetric cipher uses a public secret key *k* to encode and decode information. A symmetric encryption system contains of three original functions:

  1. **KeyGen$_{SE}$(1$^{\lambda}$):** *k* is the key generation algorithm that generates *k* using security parameter 1^{λ}.
  2. **Enc$_{SE}$ (*k*,*M*):** *C* is the symmetric encryption algorithm that takes the secret and message *M* and then outputs the cipher text *C*.
  3. **Dec$_{SE}$ (*k*, *C*):** *M* is the symmetric decryption algorithm that takes the secret cipher text *C* and then outputs the original message *M*.

- **Convergent Encryption Algorithm:** Merging cipher offers data confidentiality in deduplication. A user (or data owner) derives a convergent key from each original data copy and encrypts the data copy with the merging key. In adding, the user also develops a label for the data photocopy, such that the label will be used to identify replacement. Here, we assume that the tag perfection assets hold, i.e., if two data copies are the similar, then their labels are the similar. To discover duplicates, the user first sends the label to the server adjacent to checkered if the equal copy has been previously put in storage. Note that both the merged key and the label are individually resultant and the label cannot be used to deduce the convergent key and cooperation data privacy. Both the encoded data copy and its equivalent label will be stored on the server side. Formally, a merged cipher scheme can be defined with four original purposes:

1. **KeyGen$_{CE}$(M):** K is the key generation algorithm that maps a data copy M to a convergent key K.
2. **Enc$_{CE}$ (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 cipher text C.
3. **Dec$_{CE}$ (K,C):** M is the decryption algorithm that takes both the cipher text and C the convergent key K as inputs and then outputs the original data copy M.
4. **TagGen(M):** T(M) is the tag generation algorithm that maps the original data copy M and outputs a tag T(M).

In cryptography, a keyed-hash note confirmation code (HMAC) is a specific building for computing a message authentication code (MAC) involving a cryptographic hash function in grouping with a secret cryptographic key.

V. Proposed Algorithm

It may be used to frequently confirmed both the data integrity and the authentication of a message. Any hash function, like MD5 or SHA-1, used in the computation an HMAC; the resulting MAC algorithm is termed HMAC-MD5 or HMAC-SHA1 considered. The cryptographic potential of the HMAC belongs upon the cryptographic potential of the underlying hash function, the volume of its hash result, and on the volume and quality of the key.

$$HMAC(K,m) = H((K \oplus opad)|H((K \oplus ipad)|m))$$

- H is cryptographic hash purpose
- K is a secret key padded to the right with extra zeros to the input block size of the hash function, or the hash of the individual key if it's longer than that block size
- | denotes concatenation,
- Big plus denotes exclusive or (XOR)
- m is the message to be authenticated
- opad is the outer padding (one-block-long hexadecimal constant)
- ipad is the inner padding (one-block-long hexadecimal constant)

VI. Conclusion

The concept of authorized data deduplication was proposed to protect the data security by including differential privileges of users in the duplicate check. Similarly open several new deduplication constructions supporting authorized duplicate check in hybrid cloud design, in which the identical form symbols of folders are generated by the private cloud server with private keys. Security analysis proves that our schemes are secure in terms of insider and outsider attacks specified in the proposed security ideal. As a resistant of idea, implemented a prototype of authorized duplicate check scheme and conduct tested experiments on our prototype. System showed that authorized duplicate check scheme sustains minimal overhead compared to merging cypher and network transfer.

References

[1] OpenSSL Project. http://www.openssl.org/

[2] P. Anderson and L. Zhang. Fast and secure laptop backups with encrypted de-duplication. In Proc. of USENIX LISA, 2010.

[3] M. Bellare, S. Keelveedhi, and T. Ristenpart. Dupless: Serveraided encryption for deduplicated storage. In USENIX Security Symposium, 2013.

[4] M. Bellare, S. Keelveedhi, and T. Ristenpart. Message-locked encryption and secure deduplication. In EUROCRYPT, pages 296ï¿½312, 2013.

[5] M. Bellare, C. Namprempre, and G. Neven. Security proofs for identity-based identification and signature schemes. J. Cryptology, 22(1):1ï¿½61, 2009.

[6] M. Bellare and A. Palacio. Gq and schnorr identification schemes: Proofs of security against impersonation under active and concurrent attacks. In CRYPTO, pages 162, 177, 2002.

Anup V. Adke he is Engineering student of Information Technology at Brahma Valley College of Engineering And Research Institute, Nasik under University of Pune. His interest in the field of development.

Akash N. Nawale he is Engineering student of Information Technology at Brahma Valley College of Engineering And Research Institute, Nasik under University of Pune. His interest in the field of security.

Pravin N. Jadhav he is Engineering student of Information Technology at Brahma Valley College of Engineering And Research Institute, Nasik under University of pune. His interest in the field of security.

Vijayandra A. Yeole he is Engineering student of Information Technology at Brahma Valley College of Engineering And Research Instituted, Nashik Under University of Pune. His interest in the field of database administrator.

K. S. Kumavat, ME, BE Computer Engg. Was educated at Pune University. Presently she is working as Head Information Technology Department of Brahma Valley College of Engineering and Research Institute, Nasik, Maharashtra, India. She has presented papers at National and International conferences and also published papers in National and International Journals on various aspects of Computer Engineering and Networks. Her areas of interest include Computer Networks Security and Advance Database.