International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 04 Issue: 03 | Mar -2017 www.irjet.net p-ISSN: 2395-0072

AUTHORIZED DUPLICATE CHECK SCHEME

Mr. K.Ravindran[1], K.Gayathri [2], J.Geethupriya [3], M.Manoranjitham [4], V.Pavithra[5]

[1]Assistant Professor, Department of Information Technology, Valliammai Engineering College, Anna University, Chennai.

[2,3,4,5] Student Scholars, Department of Information Technology, Valliammai Engineering College , Anna University, Chennai, Tamil Nadu, India

---

Abstract - *Data Deduplication is the process of eliminating repeated data copies. It provides security and privacy concerns arise as user's sensitive data are susceptible to both inside and outside attacks. Traditional encryption, providing data confidentiality is incompatible in deduplication. It also requires different users to encrypt their data with their own keys. Thus, identical data copies of different users will lead to different ciphertext makes deduplication impossible. Convergent Encryption has been proposed to enforce the data confidentiality and makes the data deduplication possible. It encrypts/decrypts the data copy by using the convergent key. The convergent key can be obtained by computing the cryptographic hash value of the data content. After data encryption users retain the keys and the encryption operation is derived from the data content, identical data copies will generate the same convergent key and ciphertext. To protect data security, this project makes an attempt to address the problem of authorized data deduplication.*

***KeyWords*: confidentiality, convergent key, deduplication, encryption, hash value.**

1. INTRODUCTION

Authorized duplicate check scheme is a deduplication concept. Data deduplication is one of the important techniques for eliminating duplicate copies and has been widely used in cloud storage to reduce storage space. The deduplication has been widely used to reduce the amount of storage space and save bandwidth. To protect the confidentiality of data, authorized duplicate check is used. Apart from the traditional duplication system, different advantages of user are used. To reduce the computing time and response time between token request and response, file upload or download request and results. The authorized duplicate check is used in the hybrid cloud mechanism. The hybrid cloud architecture consists of both the public cloud and the private cloud. Private cloud plays an important role in the system, moreover security of the system is less, as the private cloud of the mechanism is not secured, and unauthorized access of the data results security fails. To provide more security, the private cloud is provided by multilevel authentication**.** This shows the system is secure and confidential. Example, the identical file may be saved in several different places by different users. Deduplication has abolished these duplicate copies by saving just one copy of the data and replacing the other copies with pointers that lead back to the original copy. The convergent encryption technique used to encrypt the user's data before uploading, to protect the confidentiality of the data. The authorized duplicate check is implemented with the encryption technique to provide the confidential data in hybrid cloud storage.

The objectives of this paper are to protect the data security in the duplicate check and to reduce the storage space. Advanced deduplication system – Supports authorized duplicate check and compare the storage system with the file content.

It has three sub phases: file upload, hash value generation, File download. The first phase is the authorized user upload the file. The second phase of hash value generation phase, we have used the MD5, a SHA-1 algorithm which generates the hash value for the uploaded file and then the hash value of the file is compared with the database. If the hash value is different then its uploaded otherwise its discarded. If the file is same then the root file is provided to the user. For the third phase, the user downloads the file.

The rest of this paper proceeds as follows. In section 2, exhibits the system analysis related work such as existing system problem, proposed system solution. In section 4, we propose the system design and functional design section 5. In section 6 deals with the conclusion and in the section 7 is the future enhancement.

2.SYSTEM ANALYSIS

The purpose of the System Analysis makes the brief analysis task and also to establish complete information about the concepts, behavior and other constraints such as performance measurement and system optimization. The goal of the System Analysis is to completely specify the technical details for the main concept in a concise and unambiguous manner.

2.1 Existing System

The convergent encryption technique used to encrypt the data before contracting. To better protect data security, it makes the attempt to the problem of authorized data Deduplication. Apart from traditional Deduplication systems, the differential privileges of users are further considered in duplicate check besides the data itself. It presents several new Deduplication techniques supporting authorized duplicate check in hybrid cloud architecture. Security analysis demonstrates that the scheme is secure in terms of the security model. As a proof of concept, it implements a prototype authorized duplicate check scheme. It shows authorized duplicate check scheme incurs minimal overhead compared to normal operations.

The drawbacks of the existing system are that this traditional convergent encryption will be insecure for predictable file. Convergent encryption, to support duplicate check, the key is derived from the file *F* by using the same cryptographic hash function.

2.2 Proposed System

The advanced duplication system supports the authorized duplicate check and compares the storage system. The *file* uploaded once it makes the hash value. In this way, the users cannot upload the same hash value data because it compares the whole database storage system, which means that it can prevent the privilege key sharing among users in the above straightforward construction. The user needs to send a request to the private cloud server to get a file token. Duplicate check is performed by comparing the storage system for any file. The user needs to get the file token from the private cloud server. The private cloud server checks the user's identity before issuing the corresponding file token to the user. Before uploading this file, the authorized duplicate check can be performed by the user with the public cloud. Based on the results of duplicate check, the user either uploads this file or runs Paw.

The benefits of our proposed work is the design and implementation of the new system which could protect the security of predictable message. The technique is the novel encryption key generation algorithm. For simplicity, it uses the hash functions to define the tag generation functions and convergent keys. The system also protects the data security by including differential privileges of users in the duplicate check.

3.SYSTEM DESIGN

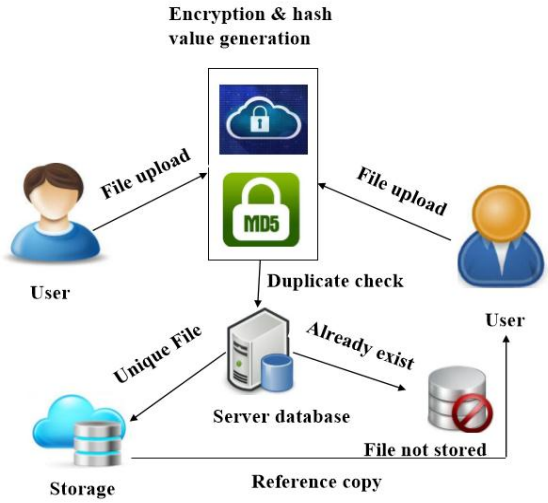

Figure 1: Architecture diagram for the proposed work

The above figure 1 shows the architecture of the proposed system. The authorized user can upload the file into the server. The uploaded file can be encrypted using the AES algorithm and the hash value for that file can be generated by using the MD5 and SHA-1. Every uploaded file is encrypted and the hash values are generated. The server contains the hash table for the encrypted files and these hash values are stored in MySQL database. The Deduplication process can be carried out by comparing the uploaded file hash value to the database. When two files are having the same hash value, then the root file is stored in the database and the newly arrived file is deleted. The reference copy of the root file is then sent to the user. Thus the duplication is avoided.

4.FUNCTIONAL DESIGN

4.1 Methodology

The methodology of this paper is that the data source given for a project is an image file. The file can be either

upload, download or delete by the user. Whenever the user wants to upload a new file, it can be encrypted using AES algorithm. The hash value for the uploaded image file can be generated by using the MD5 and SHA-1 algorithm. The RGB value for each pixel is computed. And the computed hash values are stored in the database. Only the unique files can be stored in the database. Then the user is considered as a root user and the file can be denoted as root file. If other users want to upload the same file, the reference copy of the root file can be provided to the user. Authentication can be provided by using the OTP generation process. The OTP can be generated for the user who wants to upload the file. The user can also download the image file. The encrypted image file can be decrypted by using the AES algorithm. The file deletes process can also be carried out by the user. Whenever a root file is deleted by the root user, the next user who has the reference copy.

4.2 Modules

The modules of the proposed system are registration, file upload, duplicate check, file download, file delete.

4.2.1 Registration

The user sends the mail id to the admin. The admin is responsible for generating the OTP to the user. The OTP is generated in the user mail. After generating the admin permission, the registration process can be carried out. In the registration process, the user has to specify the mail-id and the OTP. The authentication takes place, whether the mail-id and OTP are correct. If it corrects, then the user need to provide the following details such as username, mail id, gender, mobile number and product key value. After the registration, the details are stored in the database. Only the authorized user can upload/download the file.

4.2.2 File Upload

In this file upload module, the user need to choose the file that the user wants to upload. Then the AES algorithm is used for file encryption process. It encrypts the file and then generate hash values.

4.2.3 File Duplicate check

After the encryption of the file, hash value for the uploaded file can be generated by using MD5 and SHA algorithm. The user, who upload the first file referred as the root user and the file is denoted as root file. If the same file is uploaded by other user, reference copy of the root file can be provided.

4.2.4 File Download

In this file download module, the user needs to send the file download request to the admin. After receiving the response, the file download process can be carried out. Then, AES decryption algorithm is used for decrypting the file.

4.2.5 File Delete

In this file delete module, the users are allowed for deleting the files. If the root file is deleted by the user then the next user who get the reference copy, referred as root user. In such a way the process can be carried out.

5.CONCLUSION

The duplicates are being reduced to store more data and to reduce the storage space. To increase the security, we propose the multilevel authentication technique using OTP generation. As a result, to provide more security to the private cloud it uses multilevel authentication. It checks for duplication and generate the token by the private cloud before uploading the file. Likewise, the file download and delete is also performed through which storage space is effectively used.

6.FUTURE ENHANCEMENT

To further enhance, need to check deduplication for films uploading, books with large pages, videos and audios with efficient time. Although data deduplication has been studied for more than ten years, many open problems and challenges remain to be addressed, particularly as the size of digital data continues to grow exponentially and the need for long-term storage management becomes increasingly urgent. There will be more applications for deduplication, such as storage systems for tapes or shingled disks, since it will help to reduce the growing redundant data in large-scale storage systems.

REFERENCES

[1] Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P.C. Lee, and Wenjing Lou, “A Hybrid Cloud Approach for Secure Authorized Deduplication”, IEEE transactions, 2015.

[2] Jiawei Yuan and Shushing Yu, “Secure and Constant Cost Public Cloud Auditing with Deduplication”, Conference Paper, 2015.

[3] Mahir Buller Chana hip Namprempe and Gregory Neven, "Security Proofs for Identity-Based Identification and Signature Schemes", Conference Paper 2014.

[4] Wee Kong Ng and Yonggang Wen, "Private Data Deduplication Protocols in Cloud Storage, Conference Paper, 2014.

[5] Jorge Balasko, Augustin Orillia, Carlos III, Robert DI Pietro, Alessandro Sorniotti "A Tunable Proof of Ownership Scheme for Deduplication using Bloom Filters", IBM Research, 2013.

[6] John R. Douceur, Atul Adyta, William J. Bolo sky, Dan Simon, Marvin, "Reclaiming Space from Duplicate Files in a Server less Distributed File System" Theimer Microsoft Research,2013.

[7] Jingwei Li, Jin Li, Dongqing Xie and Zhang Cai, "Secure Auditing and Deduplicating Data in Cloud" IEEE Transactions, 2015.

[8] K. Saritha, S.Subasree "Analysis of Hybrid Cloud approach for Private Cloud in the De-Duplication Mechanism", IEEE Transactions, 2015.

[9] Gaurav Kakariya, Sonali Rangdale, "A Hybrid Cloud Approach for Secure Authorized Deduplication" International Journal Computer Engineering and Applications, Volume VIII, Issue I,2014.

[10] J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer, "Reclaiming space from duplicate files in a serverless distributed file system," in Proc. International Conference. Distrib. Comput. System,2002.