International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 04 | Apr 2020 www.irjet.net p-ISSN: 2395-0072

A Novel Approach Implementing Deduplication using Message Locked Encryption

Kapil Srivastava[1] , Rahul Saini[2], P. Mohamed Fathimal[3]

[1]Kapilsrivatava, Student, Dept. of Computer Science Engineering, SRM Institute of Science and Technology, Chennai, Tamilnadu, India
[2]Rahul Saini, Student, Dept. of Computer Science Engineering, SRM Institute of Science and Technology, Chennai, Tamilnadu, India
[3]P.Mohamed Fathimal, Proffessor, Dept. of Computer Science Engineering, SRM Institute of Science and Technology, Chennai, Tamilnadu, India

---

ABSTRACT: *Cloud is the most prominent and fastest-growing technology after the start of the 20th century. Since it becomes, most brilliant and intelligent service, accessible across all the domains of technology makes it vulnerable to attacks, and even some other errors like replicating of data, the efficiency of storage, and the amount of space needed. To sort the problem of duplication and storage of data, Data compression known as deduplication used on the cloud. We have implemented a system that overcomes the limitations which use Erasure Code technology, tokenization, and encryption algorithms to protect data loaded onto the cloud by the client. Our implementation sees not only one side of the problem but also another side; this is how. People always want their particulars, personals, and essential date to be perfectly safe and secured, and nobody should control and rule over except them themselves. So instead of making a third-party auditor, we have given the power to the client to choose who is the person can check and access the there data. The client can even check what happens with the data, which uploads onto the cloud, which ensures the proof of ownership principle. Here, I like all good things all together and nothing impure. It sounds perfect, but we know that perfection is a myth. Convergent encryption has some vulnerabilities. That's why we planned to implement message locked encryption, as described above.*

Keywords: erasure code technology, DES, MD5Sum, deduplication, tokenization

1. INTRODUCTION

In this article, we see the process of deduplication in the cloud. Traditional deduplication algorithms such as convergent encryption which fails to protect the data. We have implemented a technique called Message Locked Encryption; in this, we used various encryption algorithms and tokenization with Erasure code technology to protect and save the data. MD5 Sum is used to check deduplication between the files on the application.

Our paper is organized as follows. First, we formally review the present state of data storing and read performance. Second, we introduce our approach. Third, we present our experimental methodology. Finally, we highlight our conclusions and outline future steps.

2. OBJECTIVE

The objective of this project is to implement secure deduplication on the cloud with all the necessary terms mentioned terms and not just making it deduplication in all aspects but also no compromise with the security.

3. LITERATURE REVIEW

3.1 Secure Enterprise and Read Performance Enhancement in data deduplication

Description: With the tremendous growth of data in the world, the use of Cloud technology and providers are gaining more popularity since these types of services provide end to end-user benefit it is essential to provide with the best computational power they need. The paper talks about the proof of retrievability and proof of ownership concept to save the data of the user from tampering or manipulation. They used data in the form of Chunk to retrieve the data readily over the period. It helps in useful communication of data.[16].

3.2 A Hybrid Cloud Approach for secure Authorised Deduplication

Description**:** paper revolves around securing the cloud storage and adding a way of maintaining ownership of data. They used the concept of differential privileges of users to check duplicate data on the cloud. They even used a hybrid cloud architecture to check the data and save it on the cloud. The proposed idea in the paper is useful for backup storage. They showed a lock level encryption on the file .even private and public cloud is a significant helpful way to show how data is secured and stored.[3]

3.3 A Survey of Secure Data Deduplication

Description: The paper discusses the way and how to effect storage in the cloud is shown by less costly means .the article discusses the backup storage and how companies are spending a lot of money on the room .it also surveys the method of deduplication which helps the companies to save the data and money. **[2]**

4. EXISTING SYSTEM

Since the modern era of the digital world has begun, .the concern over the storage and accessing of data is also increased, which inturns gain a lot of attraction of tech genius and company working to protect it. In late 2000, after the introduction of the cloud system, the way of storing and accessing the data has improved a lot in recent times. Still, various problems also arise with traditional cloud-based encryption, such as deduplication attacks on the cloud system. Convergent encryption was introduced to protect the data and remove deduplication. As time passed, the encryption technique is disregarded since it is posed to file system attacks on the server. It also does not support the principle of POW(proof of ownership). [3]

5. PROPOSED SYSTEM

Paper discusses our proposed technique that we called message locked encryption on data. Using Erasure code technology, which results in efficient storage and encryption, we are using DES and MD5 for hashing .tokenization for efficient file retrieval. Since it envelops the message with layers of encryption and by code technology, we are data breaking into the pieces before storage, connecting it by tokens for easier retrieval and protection from attacks. To check deduplication ewe used a program called MD5Sum, which acts as a checksum to check the deduplication. We have also provided proof of ownership by giving the client the power to choose whom file to be shared and can check what changes have done to file over time.

6. SYSTEM ARCHITECTURE

The project of ours has three components.

6.1 The client

6.2 The Cloud

6.3 The Recipient

6.1 THE CLIENT

A Client is a person who is the one who uploads data onto the cloud. He is the one who has the privilege to choose the person whom the data is to receive.

6.2 THE CLOUD

The cloud is the basis of implementing all the functionality in the project. All the deduplication related process occur inside the cloud. People can access after the specific people in an enterprise designate them.

6.3 THE RECIPIENT

A Recipient is a person who is selected by user access the client data.

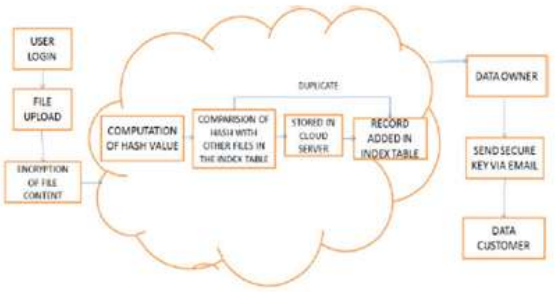

Fig 1.Architecture Diagram

7. PRELIMINARIES

7.1 SQLYOG
7.2 Erasure Code Technology
7.3 DES(data encryption Standard)
7.4 MD5Sum(checksum)
7.5 JDK(java development kit)
7.6 SMTP(mail transfer protocol)
7.7 Tokenization
7.8 Web Browser
7.9 Glassfish

7.1 SqlYog

SQLyog is a GUI tool for the RDBMS MySQL. It features Visual Schema Designer, Visual Query Builder, Query Formatter, Connectivity options for Direct client/server using MySQL API (SSL supported), and the full character set/Unicode support. [14].

7.2 ERASURE CODE TECHNOLOGY

Erasure codes are a way of data protection in which the data brakes into fragments, expanded and encoded with redundant data, and stored across a set of different locations in the storage media. This technology is used instead of RAID in various applications and current use cases in Cloud Storage [9][7].

7.3 MD5SUM(CHECKSUM)

MD5 is program that calculates and checks md5 128bit hashes. It is to verify the integrity of files, as virtually any change to a file will cause its MD5 hash to change. Most commonly, md5sum is used to confirm that a file has not changed as a result of a faulty file transfer, or non-malicious meddling. The md5sum program is implemented in most Unix-like operating systems or compatibility layers, such as Cygwin. [5][6].

7.4 DATA ENCRYPTION STANDARD

The DES is most profound and one of the most important algorithms used in encryption nowadays. The programmers use this algorithm in various ways. It is beneficial to use since it uses the plain text of size 64 bit and generates block - ciphertext of the same size [18].

7.5 JDK(JAVA DEVELOPMENT KIT)

JDK is a software used for developing java applications and applets.

7.6 SMTP

These are the security guideline used to allow the software to send electronic mail on the internet based on email addresses.

7.7 TOKENIZATION

Tokenization is a technique to provide the token to a given data to secure the given data and access it with the help of the token.[17]

7.8 WEB BROWSER

The implementation of ours is going to be in the form of a web application. So a browser is necessary, although IE-11 is preferred specifically.

7.9 GLASSFISH

This web application is to runs on the localhost with the help of glassfish. It is an application server that helps in the smooth running of JAVA EE applications.[13]

8. MODULES

8.1 DEPLOYMENT OF APPLICATION

The first step is the compiling of application and then their implementation. The directory is accessed using a tomcat server and built-in NetBeans. And can accessed on the browser.

Fig 2(a). First module

Fig 2(b). Setting up the application

Fig 3. The User Interface of the Web Application

Fig 4. Uploading of file

8.2 DEDUPLICATION MODULE

This module is heart of the application in this module.in this module we take the file as the input split them into parts using erasure code technology. we also provide

indivisual token to each splited file part using tokenization and hash each token using MD5 algorithim and store the splited file and hashes for further use .keyword is also further check the duplicacy of file

8.3 ENCRYPTION MODULE

This module deals with the encryption of file .the encrypts the file using DES algorithim

Fig 5. The Recipient can download the file.

Fig 6. Auditing the data by the client

8.4 AUDITING MODULE

This module is used to audit the file by the client (uploader) this module also ensures proof of ownership as the person to whom keyword is shared.is only allowed to access the file . It can confirm the information provided and the changes happened is right or wrong.

9. CONCLUSIONS

In this paper, the notion of effectively storing and retrieval of data in the cloud with deduplication proposed we used Erasure code technology to save the data and tokenization with some encryption algorithm to create the layer of security around the data to preserve it.

We gave power to the client to check what changes occurred on the data after uploading the data onto the cloud. It can also confirm whether the changes occurred were right or wrong in the file. On the application, we have a portal that allows the user to check the changes that occurred to file. At the back, when the receiver uploads the same file again, the server verifies it against the original hash created of file for the changes and deduplication .the result is shown to the client onto the portal.

REFERENCES

[1] A Secure Data Deduplication Scheme for Cloud Storage Jan Stanek∗ , Alessandro Sorniotti†, Elli Androulaki†, and Lukas Kencl∗Tech Rep. IBM Research, Zurich, ZUR 1308-022,2013

[2] A Survey of Secure Data Deduplication Riddhi Movaliya and Harshal Shah. Article: A Survey of Secure Data Deduplication. International Journal of Computer Applications 138(11):6-8, March 2016. Published by Foundation of Computer Science (FCS), N.Y., USA.

[3] A hybrid Cloud Approach for secure Authorised Deduplication IEEE Transactions on Parallel and Distributed Systems (Volume: 26, Issue: 5, May 1, 2015)

[4] Secure Evaluation and Prevention of Duplicate Data in Cloud A. Aarthi Sowmiya, A. Ananda Kumari, IJIRST – International Journal for Innovative Research in Science & Technology| Volume 2 | Issue 11 | April 2016 ISSN (online): 2349-6010

[5] AES and MD5 based secure authentication in cloud computing Shefali Ojha, Vikram Rajput

2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics, and Cloud) (I-SMAC)

[6] Cloud Application Security Based on Enhanced MD5 Algorithm Nalluri Sravani, Ramasubbareddy, SomulaAND Kannayaram, G Journal of Computational and Theoretical Nanoscience, Volume 16, Numbers 5-6, May 2019, pp. 2022-2027(6)

[7] Erasure Coding in Windows Azure Storage Cheng Huang, Huseyin Simitci, Yikang Xu, Aaron Ogus, Brad Calder, Parikshit Gopalan, Jin Li, and Sergey Yekhanin Microsoft Corporation

[8] Role-Based Access Controls reprint from 15th National Computer Security Conference (1992) Baltimore, Oct 13-16, 1992. pp. 554 – 563 David F. Ferraiolo and D. Richard Kuhn

[9] Erasure Coding for Cloud Storage Systems: A Survey Jun Li and Baochun Li, TSINGHUA SCIENCE AND TECHNOLOGY ISSNll1007-0214ll06/11llpp259-272 Volume 18, Number 3, June 2013

[10] Review On Data Deduplication In Cloud Computing International Journal of Advance Engineering and Research DevelopmentVolume 4, Issue 11, November - 2017

[11] Ensuring security for Fixed Block Level Deduplication in Cloud Backup Akhila K, Amal Ganesh, Sunitha C, IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661,p-ISSN: 2278-8727 PP 06-12

[12] Analyzing Data Security for Cloud Computing Using Cryptographic Algorithms Gurpreet Kaur, Manish Mahajan, Gurpreet Kaur et al. Int. Journal of Engineering Research and Applications www.ijera.com ISSN: 2248-9622, Vol. 3, Issue 5, Sep-Oct 2013, pp.782-786

[13] Glassfish available at https://javaee.github.io/glassfish/

[14] SQLYOG Available at https://en.wikipedia.org/wiki/SQLyog

[15] G. Ateniese, R. Di Pietro, L. V. Mancini, and G. Tsudik. Scalable and efficient provable data possession.

[16] Secure Enterprise and Read Performance Enhancement in data deduplication S.Ushaharani K.Kungumaraj International journal of recent technology and engineering (irjcet)

[17]about tokenization :https://en.wikipedia.org/wiki/Tokenization_(data_security).