Avoiding De-duplication on Cloud using Encryption Standards

- **Govind Joshi, Akshay Kulkarni, Pandu Kokare, Baliram Rajwad, Prof .Savita Kumbhare.**

Zeal College of Engineering and Research ,Narhe, Pune

college details

Abstract:

Cloud computing offers a new way of service provision by rearranging various sources over the Internet. The most important and popular cloud service is data storage. In order to secure the privacy of data holders, data are usually stored in cloud in an encrypted form. so, encrypted data shows new summons for cloud data de-duplication, which becomes too bad for big data storage and processing that data on cloud. Traditional de-duplication systems do not work on encrypted data. Present solutions of encrypted data de-duplication suffer from security weakness. They don't flexibly support data access control and revocation. Therefore, many of them can be deployed in practice. In this paper, we propose a system to de-duplicate encrypted data stored in cloud based on ownership challenge to the data and proxy re-encryption. It combines cloud data de-duplication with access rights. We find out its performance based on extensive analysis and computer simulations. The result shows the superior efficiency and effectiveness of the system for potential practical deployment, especially for big data de-duplication in cloud storage.

Keywords:

Access control, big data, cloud computing, data de-duplication, proxy re-encryption

Introduction

In the recent years the data is generating day by day. CLOUD computing offers a new way of Information Technology services by rearranging various resources (e.g., storage, computing) and providing them to users based on their demands. Cloud computing provides a big resource pool by linking network resources together. It has desirable properties, such as scalability, elasticity, fault-tolerance, and pay-per-use. Thus, it has become a promising service platform. The most important and popular cloud service is data storage service. Cloud users upload personal or confidential data to the data centre of a Cloud Service

Provider (CSP) and allow it to maintain these data. Since intrusions and attacks towards sensitive data at CSP are not avoidable, it is prudent to assume that CSP cannot be fully trusted by cloud users. Moreover, the loss of control over their own personal data leads to high data security risks, especially data privacy leakages. Due to the rapid development of data mining and other analysis technologies, the privacy issue becomes serious. Hence, a good practice is to only outsource encrypted data to the cloud in order to ensure data security and user privacy. But the same or different users may upload duplicated data in encrypted form to CSP, especially for scenarios where data are shared among many users.

Problem Definition:

We propose a scheme to de-duplicate encrypted data at CSP by applying PRE to issue keys to different authorized data holders based on data ownership challenge. It is applicable in scenarios where data holders are not available for de-duplication control. Therefore, few of them can be readily deployed in practice. In the existing system there was a problem to deal with encrypted data. It was failing when we try to store the encrypted data on cloud.

Objectives

The main objective of the project is to de-duplicate the data using various encryption standards. And to save your data in encrypted format. And give access to every user without leaking the main private key.

System design:

System Architecture

Proposed system we are going to develop cloud based system where our data is saved on cloud. As there is privacy and authentication issue there is need to save our data securely. For that we are using different algorithms of cryptography. Here our system helps us to de-duplicate our data so that the space on cloud can be saved and the energy and optimization also get saved. And our aim is to reduce time complexity of our project. In the following diagram there is user and he uploads the file on cloud. AP is nothing but the admin which looks on users files. CSP is cloud provider which gives you the data storage services of cloud. De-duplication is checked on encrypted data. The encrypted data is saved on cloud. And user will get the file by performing decryption.

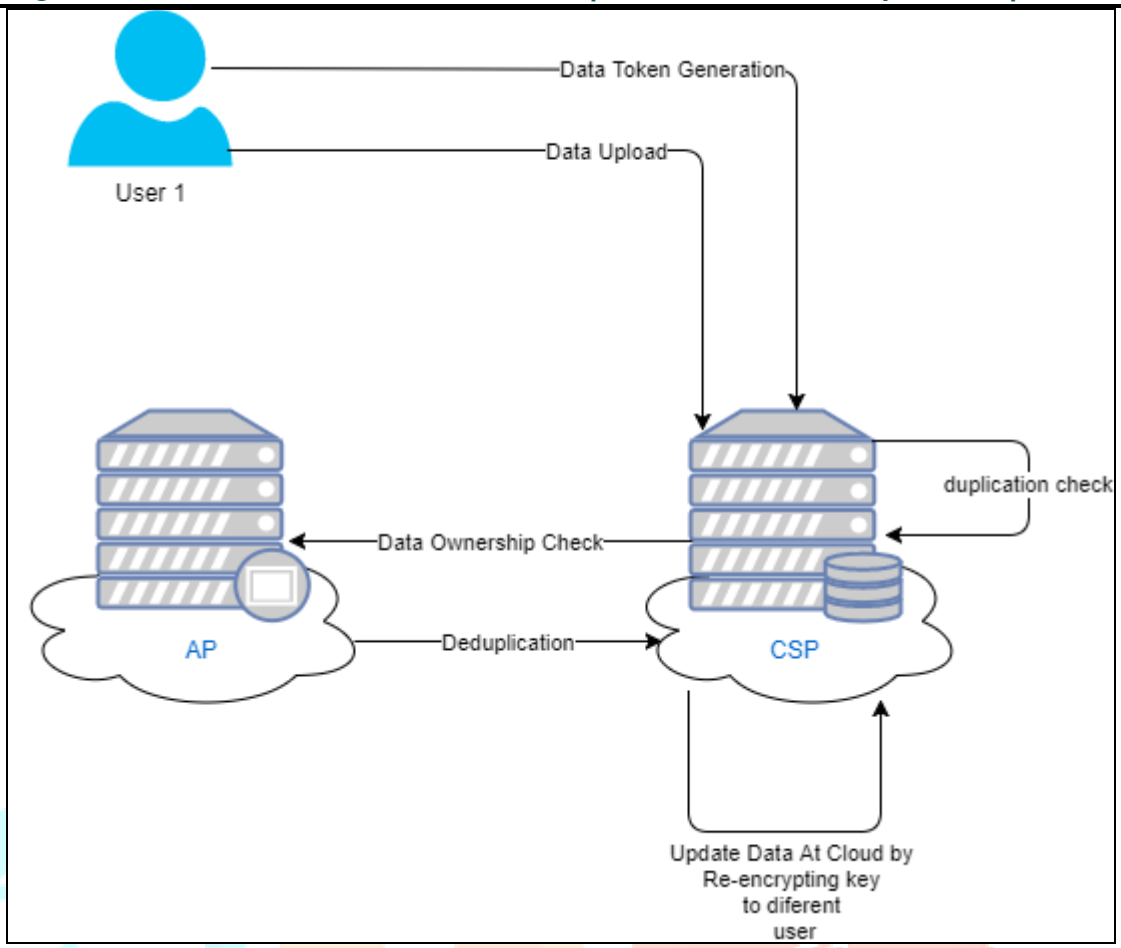

Fig 1: System Overview

Related Works

1) **DupLESS Server-Aided Encryption for De-duplicated Storage.**

Year: 2013

Author Name:

- Mihir Bellare
- Sriram Keelveedhi
- Thomas Ristenpart.

Description: This paper implement a new system called DupLESS (Duplicateless Encryption for Simple Storage) that provides a more secure, easily-deployed solution for encryption that supports de-duplication. In DupLESS, a group of affiliated clients (e.g., company employees) encrypt their data with the aid of a key server (KS) that is separate from the SS.

Limitations: This system only for security and unpredictable data may be a limitation for, and threat to, user privacy this paper shows unpredictable data. and this is very difficult to access data.

2) **ClouDedup: Secure De-duplication with Encrypted Data for Cloud Storage**

Year: 2013

Author Name:

- Pasquale Puzio
- Refik Molva
- Melek Onen
- Sergio Loureiro.

Description: The new ClouDedup system in order to implement the key management for each block together with the actual de-duplication operation. This system additional HSM can be implemented by taking advantage of Amazon CloudHSM which provides secure, durable, reliable, replicable and tamper-resistant key storage.

Limitations: This system finding possible optimizations in terms of bandwidth, storage space and computation.

3) **A Policy-based De-duplication Mechanism for Securing Cloud Storage**

Year: 2015

Author Name :

- Zhen-Yu Wang1
- Yang Lu1
- Guo-Zi Sun

Description: The new policy-based de-duplication proxy scheme using the security proxy and random storage , which separate storage services and security services to ensure the security of user data and improve the system efficiency at the same time.

Limitations: This paper we not consider duplicated data management (e.g., deletion and owner management) and did not evaluate scheme performance.

4)Reducing impact of data fragmentation caused by in-line de-duplication

Year**:** 2012

Author Name:

- Michal Kaczmarczyk
- Marcin Barczynski
- Wojciech Kilian

- Cezary Dubnicki

Description: This paper focused on inter-version duplication and proposed Context-Based Rewriting (CBR) to improve the restore performance for latest backups by shifting fragmentation to older backups.

Limitations: The full backup of only one file system is saved every week to a system with backward pointing de-duplication.

5) DeDu: Building a De-duplication Storage System over Cloud Computing.

Year: 2011

Author Name**:**

- Zhe SUN
- Jun SHEN
- Jianming YONG.

Description: This paper presents a de-duplication storage system over cloud computing. Our de-duplication storage system consists of two major components, a front-end de-duplication application and Hadoop Distributed File System.

Limitations**:** We developed efficient de-duplication system, but it cannot handle Encrypted data.

6) A Verifiable Data De-duplication Scheme in Cloud Computing.

Year: 2014

Author Name:

- Zhaocong Wen
- Jinman Luo
- Huajun Chen
- Jiaxiao Meng
- Xuan Li♯ and Jin Li.

Description: This system presents image de-duplication scheme adopts two servers to achieve Verifiability of de-duplication.

Limitations: This system is not flexible to support data access control by data holders, especially for data revocation process, since it is impossible for data holders to generate the same new key for data re-encryption.

7) Survey and Classification of Storage De-duplication Systems.

Year: 2014

Author Name:

- JO˜ AO PAULO
- JOS´E PEREIRA.

Description: This paper is offline de-duplication systems. Then provide reliability, security and privacy should be taken into considerations when designing a de-duplication system.

Limitations: This paper we not consider duplicated data management (e.g., deletion and owner management) and did not evaluate scheme performance.

8)A Hybrid Cloud Approach for Secure Authorized Deduplication.

Year: 2013

Author Name **:**

- Jin Li
- Yan Kit Li
- Xiaofeng Chen
- Patrick P. C. Lee,
- Wenjing Lou

Description**:** This paper is also several implementation of convergent of different convergent encryption variants for secure de-duplication. This system provides reliability, security and privacy with sound performance.

Limitations: This system cannot flexibly support data access control and revocation at the same time.

9)Efficient Hybrid Inline and Out-of-Line Deduplication for Backup Storage.

Year: 2015

Author Name:

- YAN-KIT LI
- MIN XU
- CHUN-HO NG
- PATRICK P. C. LEE

Description: This paper we design and implement *RevDedup*, an efficient hybrid inline and out-of-line de-duplication system for backup storage.

Limitations: This is high performance in essential operations of deduplication backup storage systems, including backup, restore, and deletion, while maintaining high storage efficiency.

10) Improving Restore Speed for Backup Systems that Use Inline Chunk-Based De-duplication

Year: 2013

Author Name:

- Mark Lillibridge
- Kave Eshghi
- Deepavali Bhagwat

Description: This paper improves the restore performance for latest backups by shifting fragmentation to older backups.

Limitations: We developed to forfeit de-duplication to reduce the chunk fragmentation by container capping. In our previous work we developed using PRE for cloud data de-duplication.

Tools Used

- **Software Requirement:**
  - Operating System : windows 8 and above.
  - Application Server : Tomcat5.0/6.X
  - Language : Java
  - Front End : HTML, JSP
  - Database : MySQL

- **Hardware Requirement:**
  - Processor : Intel i3/i7/i5
  - RAM : 4 GB (min)
  - Hard Disk : 20 G/B(min)

Statistical Technique Used

We have developed Login and Registration which manages the user profiles (User and Admin), so that the users can upload the file on cloud in the encrypted format and at the time of file download you will get your file by decrypting the file.

Algorithm

- **Data Encryption Standard(DES)**: This Algorithm is used to encrypt your data from the file so that your data will not be hacked by third party.
- **Elliptic Curve Cryptography(ECC)**: In our system, we have used ECC to generate a specific key for different files.
- **Proxy Re-encryption(PRE)**: In cryptography PRE is used to generate different keys for different users to access single file. So that key of one user will not get leaked.

.DES:

```
Cipher(byte in[16], byte out[16], key_array round_key[Nr+1])
begin
byte state[16];
state = in;
AddRoundKey(state, round_key[0]);
for i = 1 to Nr-1 stepsize 1 do
SubBytes(state);
ShiftRows(state);
MixColumns(state);
AddRoundKey(state, round_key[i]);
end for
SubBytes(state);
ShiftRows(state);
AddRoundKey(state, round_key[Nr]);
end
```

ECC:

1. add some extra data to the end of the input

```
        set the initial sha-1 values
            for each 64-byte chunk do
              extend the chunk to 320 bytes of data
              perform first set of operations on chunk[i] (x20)
              perform second set of operations on chunk[i] (x20)
              perform third set of operations on chunk[i] (x20)
```

perform fourth set of operations on chunk[i] (x20)

end

return value as a key

PRE :

Input : file

Output : Decrypted file download

Step 1: Take file as a input

Step 2 : Generate key

Step 3 : get recepeint id

Step 4:send key to mail

Step 5: exit

Our Approach:

The system will work in three operating modes:

1. User :

The user is the holder and owner of the data file. The data will be uploaded by the user, this file will be saved on cloud in encrypted format.

2. Admin :

The admin is a intermediate between user and cloud. Here you can see which uploaded which file.

3. View Files :

You can see list of files here for downloading.

4. Delete Files :

Here you can delete a file.

Experiment Result:

This system will collect the files by downloading by user this files will be decrypted files. User will get a key on user's mail. This key will be different for different users so that main key will not get hacked.

Future scope:

In future, we will develop a flexible system which can work with real time data. In the future we will consider mp3, mp4 data and pdf files also.

Acknowledgment: (optional)

It gives us great pleasure in presenting the preliminary project report on 'Avoiding De-duplication on Cloud Encryption Standards'.

I would like to take this opportunity to thank my internal guide Prof. S. A. Kumbhare. For giving me all the help and guidance I needed I am really grateful to them for their kind support. Their valuable suggestions were very helpful.

I am also grateful to Prof. ……., Head of Computer Engineering Department, …….., for his indispensable support and suggestions.

Govind Joshi

Akshay Kulkarni

Pandu Kokare

Baliram Rajwad

(B.E. Computer Engineering).

Conclusion:

In the recent days the data storage with de-duplication is very important topic. In the existing system the data storage on encrypted data including de-duplication is very hard to achieve. So here there is need to store encrypted data on cloud with data security. In this paper we proposed a model to deal with the data storage with encryption that will de-duplicate the data. Here we are going to manage the encryption of the data with ownership challenge. And proxy re-encryption. Our scheme can be flexibly support the data sharing among different users. Encrypted data can be securely accessed without leaking the original data and its key. And one more thing is encrypted data will be accessible to the authorized data holder. Here we are going to use symmetric keys for data decryption which will send to the authorized user's mail id. Performance analysis shows that our scheme is suitable to store the big data on cloud. Future work includes optimizing our design and implementation for practical deployment and studying verifiable computation to ensure that CSP behaves as expected in de-duplication management.

Reference:

[1] M. Bellare, S. Keelveedhi, and T. Ristenpart, "DupLESS: Server aided encryption for deduplicated storage," in Proc. 22nd USENIX Conf. Secur., 2013, pp. 179–194.

[2] Dropbox, A file-storage and sharing service. (2016). [Online]. Available: http://www.dropbox.com

[3] Google Drive. (2016). [Online]. Available: http://drive.google.com

[4] Mozy, Mozy: A File-storage and Sharing Service. (2016). [Online]. Available: http://mozy.com/

[5] 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. IEEE Int. Conf. Distrib. Comput. Syst., 2002, pp. 617–624, doi:10.1109/ICDCS.2002.1022312.

[6] G. Wallace, et al., "Characteristics of backup workloads in production systems," in Proc. USENIX Conf. File Storage Technol., 2012, pp. 1–16.

[7] Z. O. Wilcox, "Convergent encryption reconsidered," 2011.n[Online]. Available:http://www.mailarchive.com/cryptography@metzdowd.com/msg08949.html

[8] G. Ateniese, K. Fu, M. Green, and S. Hohenberger, "Improved proxy re-encryption schemes with applications to secure distributed storage," ACM Trans. Inform. Syst. Secur., vol. 9, no. 1, pp. 1– 30, 2006, doi:10.1145/1127345.1127346.

[9] Opendedup. (2016). [Online]. Available: http://opendedup.org/

[10] D. T. Meyer and W. J Bolosky, "A study of practical deduplication," ACM Trans. Storage, vol. 7, no. 4, pp. 1–20, 2012, doi:10.1145/2078861.2078864.