IJSTE - International Journal of Science Technology & Engineering | Volume 3 | Issue 01 | July 2016
ISSN (online): 2349-784X

A Novel Mechanism for Secure and Authorized Deduplicaton using Hybrid Cloud

Shabnam Siddiqui
M. E. Student
Department of Computer Engineering
Dhole Patil College of Engineering, Pune, India

Arati Dandavate
Head of Dept.
Department of Computer Engineering
Dhole Patil College of Engineering, Pune, India

Abstract

Nowadays Cloud Computing is one of the greatest and important technologies. Cloud computing mainly means storing and using data and program over the internet instead of using it from computer hard drive. As the cloud is mainly used for storage purpose, the one of the problem is space utilizations. Sometime same data gets uploaded twice and hence the cost of data storage gets increased. To overcome data storage issue data deduplication becomes necessary. Data deduplication mainly eliminates duplicate and redundant data and thereby save the bandwidth over the cloud. To encrypt the data before it is send, a highly secure encryption technique has been proposed. The proposed system essentially address the problem of authorized and data deduplication. It mainly uses hybrid cloud architecture to support authorize deduplication. The proposed system combines both the approaches of securing the data and deduplicating the data in correspondence to each other.

Keywords: Authorization, Data Security, Privilege, DE Duplication, Credentials, Cloud

I. INTRODUCTION

Cloud Computing refers to the on demand computing which is a kind of internet based computing where shared resources, information and data is provided to the computer on demand. Hybrid cloud is mainly mixture of two clouds. A high degree of fault tolerance is obtained by using the hybrid cloud architecture. Hybrid cloud mainly work at infrastructure and application level. Cloud can offer the possibility of storing the files and accessing, storing and retrieving them from any web-enabled interface. Using cloud at any time and place give user the high availability, speed, scalability and security for your environment. Organization need to pay for the amount of storage they are actually consuming. Data is stored in virtualized pools of storage hosted by third party mostly.

Private cloud is cloud infrastructure operated solely for a single organization, which is managed internally or by a third-party or externally. Private cloud mainly provides computing power as service in virtualized environment. It mainly uses pool of physical computing resources. The word cloud is used as a metaphor for “the internet”. The cloud computing mainly means "a type of Internet-based computing" where different services — such as servers, storage and applications are provided to an organization's computers and devices through the Internet. In such an authorized de-duplication system, during system initialization each user is provided a set of privilege. To perform the duplicate check a set of privileges are linked to each file which indicate the kind of user are allowed to access the file. Before submitting his duplicate check request or some file, the user needs to take this le and his own privileges as inputs. The user is able to find a duplicate for this le if and only if there is a copy of this file and a matched privilege stored in cloud [2]. A traditional deduplication system mainly uses convergent encryption, which provides confidentiality to some extent. But it does not support the duplicate check with differential privileges. It was difficult to realize both deduplication and different differential authorization at the same time. Cloud-based services are mainly used for businesses with growing or fluctuating bandwidth demands. If needs increases it is to scale to the cloud.

II. LITERATURE REVIEW

Many people store large amount of personal and corporate data on laptops or home computers. These often lead to hardware failure. This paper mainly describe an algorithm which mainly focused on data which is common between users. The drawback of this algorithm is it is mainly used for laptop backup.

P. Anderson mainly defined an algorithm which is used mainly for Fast and secures laptop backups with encrypted deduplication[1]. The deduplication was performed on backup data that was created for data stored on. The drawback with this algorithm was it can be used with laptop backup data only.

T. Schneider mainly defined an algorithm which is used mainly for secure cloud computing[2]. They have defined twin clouds where one cloud is used for storing the data and other cloud is used for analysing the queries. Ferraiolo describes an algorithm which mainly includes mechanism for convergent encryption which enables duplicate files to coalesce into the space of a single file and Self- Arranging and location information in a decentralized manner.

Naveen mainly focused an algorithm where the security proofs or attacks based on identity-based identification for different files and signature schemes are defined either explicitly or implicitly in existing literature. Underlying are the framework that on

the one side enables the modular security analyses and in other side[3] helps to explain how these schemes are derived thereby helping to understand, simplify the previous work. The main purpose of cloud computing is to provide cloud services and user need to pay only for the resources used. The existing system mainly used to perform file level deduplication which was not completely authorized.

III. PROBLEM DEFINITION

As cloud provide different services like platform as service, software as service, Infrastructure as service the problem like cloud storage arises. The main problem of cloud storage services is the management of the high volume of the increasing data. Beside from traditional duplication systems, the data itself and the differential privileges assigned to the users are further considered. Also presented several new de-duplication constructions which support authorized duplicate check in hybrid cloud architecture. In support with the duplicate check, the system should provide authorize duplicate check. In hybrid cloud architecture, several new de-duplication constructions are presented. Security analysis demonstrates that our scheme is secure. We implement a proposed authorized approach for duplicate check which will take minimal overhead compare to other operations.

IV. IMPLEMENTATION DETAILS

A. Proposed System Architecture

In proposed system our aim is to perform authorize deduplication and maintaining the file security. Also to prevent and protect the confidentiality of sensitive and important data while supporting deduplication over clouds, the novel and highly secure encryption technique has been proposed where the data is encrypt before it is outsource. In this system, designing of the different modules is undertaken.

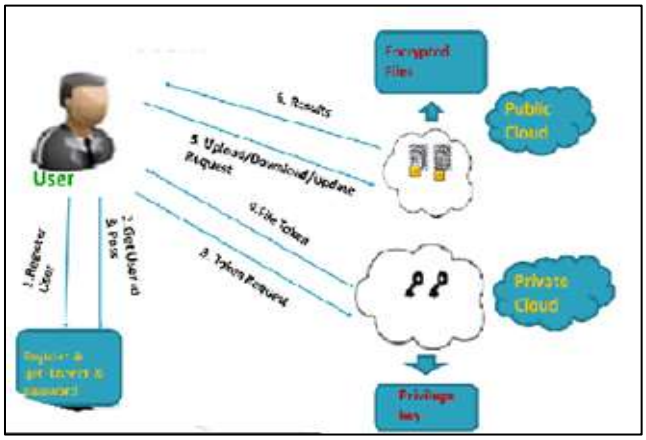

Fig. 1: System Architecture

Whenever user wants to upload a file on a cloud he needs to register himself first into the cloud. Once the user is registers successfully, he will received user id and password. To make the system more authorize, user needs to authorize himself before any file is upload or download.

For authorization user login password and OTP are used in combination. The private cloud will keep the data of the file token and the respective privileges. The public cloud will be used to store the encrypted files. The files are encrypted so that the external agents are not able to access them directly.

B. Mathematical Model

1) User Authentication
Set (C) = c0, c1, c2, c3, c4
C0= Get User Id
C1=Get Cloud Id.
C2=Get Data Owner Info
C3=get the User Privilege Information
C4= Get key from hash table
2) Data De Duplication
Set (T) = c1, c2, d0, d1, d2
d0=Get Data Filename.
d1=Data accessing userid.
d2=Get Cloud id

3) Encryption Module
Set (E) = e0, e1, e2, c1, c2
e0=get le to be encrypted
e1=get public key for encryption
e2=encryption of data
4) Token Generation Module
Set (G) = c0, c3, g0, g1, g2
g0= get user request
g1= map user privilege
g2= generate le token
5) Service Module
Set (S) = s0, s1, s2, c1, c2
s0=get user id and le request
s1=get data to be uploaded or downloaded
s2=provide service
6) Union and Intersection of Project
Set (P) = c0, c1, c2, c3
Set (t) = c1, c2, c3, d0
(C U T)=c0,c1,c2,c3,c4,d0,d1,d2
(C intersection T)= c1,c2
(C intersection E) = c0,c1

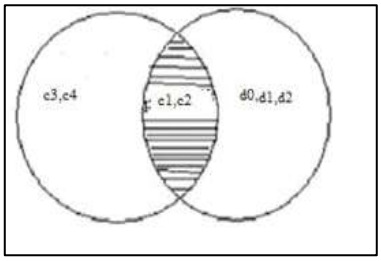

Fig. 2: Set (C intersection T)

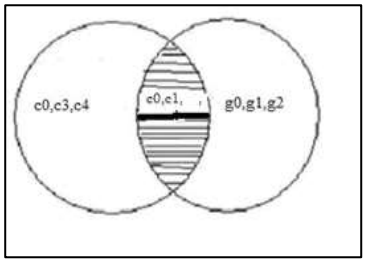

Fig. 3: Set (C intersection T)

V. RESULT

The project is going to mainly perform the deduplication check in authorized manner. Below are the some modules which are developed in this project:

A. Home Page

Home page is the first page which user will see. It mainly contains the section from where user can navigate to Home, admin, private cloud, user section.

Fig. 4: Home page

B. User Registration

User registration module helps user to register themselves. Through user registration page user will register for duplicate check.

Fig. 5: User registration

C. *Authorization Module*

This module will authorize the user using his login password and OTP.

Fig. 6: Authorization Module

VI. CONCLUSION

Thus the proposed system is capable of securely authenticating and checking the deduplication of the data over cloud by both filename as well as block level deduplication thereby promoting the advanced OTP and LTP technique for user authentication, and secure encryption and deduplication mechanism with multiple level checking for duplicate data. The proposed system will hence be used in all the cloud architectures thereby reducing the storage over the cloud and decreasing the redundant data storages over cloud.

VII. FUTURE SCOPE

In this project we are mainly working on .txt and .doc files. We are performing deduplication check on .txt and .doc files. In future we can work on .xlsx and .pdf files while performing deduplication while uploading them.

ACKNOWLEDGMENT

Special thanks go to authors Jin Li, Yan Kit Li, Xiaofeng Chen for valuable existing work in this area. I am thankful to my Guide and Head of Department Prof. Aarti Dandavate ,PG Co-ordinator Prof. Dange Varsha,Contributed to this paper for their valuable comments and sharing their knowledge and idea.

REFERENCES

[1] P. Anderson and L. Zhang. Fast and secure laptop backups with encrypted deduplication. In Proc. of USENIX LISA, 2010.
[2] S. Bogie, S. Nurnberg, A. Sadeghi, and T. Schneider. Twin clouds: An architecture for secure cloud computing. In Workshop on Cryptography and Security in Clouds (WCSC 2011), 2011.
[3] M. Bellaire, S. Keelveedhi, and T. Ristenpart. Message-locked encryption and secure de-duplication. In UROCRYPT, pages 296– 312, 2013.
[4] M. Bellaire, C. Namprempre, and G. Naveen. Security proofs for identity-based identification and signature schemes. J. Cryptology, 22(1):1–61, 2009.
[5] M. Bellaire and A. Palacio. Go and scour identification schemes: roofs of security against impersonation under active and concurrent attacks. In CRYPTO, pages 162–177, 2002.
[6] M. Bellaire, S. Keelveedhi, and T. Ristenpart. Duplets: Serve raided encryption for Deduplicated storage. In USENIX Security Symposium, 2013.
[7] J. R. Douceur, A. Adyta, W. J. Bolo sky, D. Simon, and M. Theiler. Reclaiming space from duplicate files in a server less distributed file system. In ICDCS, pages 617–624, 2002.
[8] Ferraiolo and R. Kuhn. Role-based access controls. In 15th NIST-NCSC National Computer Security Conf., 1992.
[9] GNU Libmicrohttpd. http://www.gnu.org/software/libmicrohttpd/.
[10] S. Halevy, D. Henrik, B. Pinks, and A. Shulman-Pele. Proofs of ownership in remote storage systems. In Y. Chen, G. Danes, and V. Shmatikov, editors, ACM Conference on Computer and Communications Security, pages 491–500. ACM, 2011.