Secure Authorized Deduplication on Hybrid Cloud

Taranpreet Bhatti[1], Ashish James[2], Siddhi Narvekar[3] , Prof. Varsha Wangikar[4]

[1]K.C. College of Engineering and Management Studies & Research, Mumbai University
[2]K.C. College of Engineering and Management Studies & Research, Mumbai University
[3]K.C. College of Engineering and Management Studies & Research, Mumbai University
[4]Professor, Dept. of Information Technology, K.C. College of Engineering and Management Studies & Research, Maharashtra, India.

---***---

Abstract -*Data Deduplication is one of the best data compression techniques. The paper A Hybrid Cloud Approach for Secure Authorized Deduplication implements both File Level and Block Level Deduplication on Hybrid Cloud so as to avoid repetitive storage of same data uploaded by the any user .Data here is encrypted before deduplication using AES (Advanced Encryption Standard) algorithm. Users get registered to the cloud in order to store their data and are allowed to perform actions like uploading and downloading files only after he get permission from the authorized user of the requested domain. The Security of the file is ensured by providing OTP's to the user that checks to see if the user is authenticated. Proof of ownership is used in order to trace the MAC and IP address of unauthorized user or the hacker*.

Key Words:Hybrid cloud, Deduplication, Security.

1. INTRODUCTION

Rapid growth of data and its storage is major problem that most of the organizations are facing today. Storage issue can be easily handled using cloud, but storing large and confidential data on cloud is not that easy. Using cloud can not only be expensive but also could be a threat to the confidentiality and security of the data. The basic objective of A Hybrid Cloud Approach for Secure Authorized Deduplication is that is uses Deduplication that only eliminates duplicate copies of same data uploaded thus saving storage space and bandwidth and reducing the budget. The designed system also ensures security of the data by passing the user to various authentication and authorization checks in order to access the data.

2. EXISTING SYSTEMS

Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee and Wenjing Louproposed proposed the paper "A Hybrid Cloud Approach for Secure Authorized Deduplication". This paperproposes hybrid cloud approach instead of directly using the hybrid cloud. The prototype introduced in the paper uses two different cloud i.e. Pubic cloud and Private cloud. The security of the cloud is ensured by providing privilege keys, tokens and convergent key. Algorithms like SHA-1, HMAC-SHA and Proof of Ownership. The paper is not only difficult to be understood but also too complex, expensive and time consuming to be implemented. Our papers offer additional security as compared to this paper by tracing and providing the MAC and the IP address of the hacker's machine

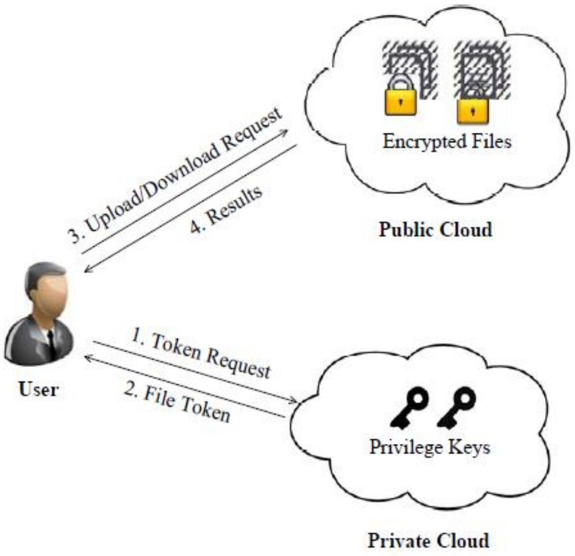


Fig – 1: Architecture for existing system

3. PROPOSED SYSTEM

To overcome the problem of complexity, to involve only one cloud, enhance security and also make the system user friendly we have tried to make the system simple and user friendly.In this paper we introduce a system that can be used by large organization in order to store their data in in less possible space, making it available to all the users and in a secured manner..

3.1 Basic Architecture

Example -1 Let us consider college systems which consist of many Departments like Administration, Examination

Control, Information Technology, Electronics etc. and students too. All the repeated data of these departments are stored in cloud thus occupying a huge space. Suppose all the department has to store a file which is common, so instead of storingN no. of copies of it only oneoriginal file can be stored on cloud using deduplication thus eliminating the repeated copies and reducing storage space. Also privileges can be assigned to file as to whom all a particular file should be visible, thus ensuring privacy and the confidentiality.

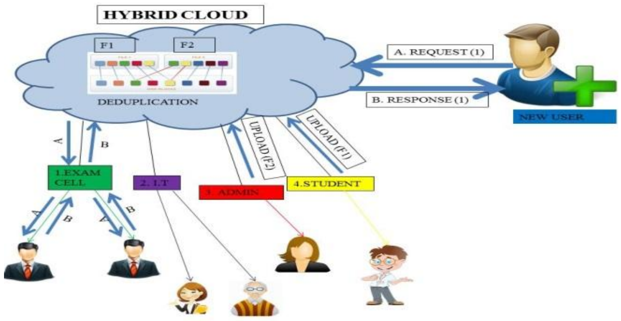


Fig – 2: Overall Architecture of the System

3.2 Design

In order to implement it we have used the following environments.

1. Bootstrap, CSS, HTML- Designing pages

2. JavaSE- 1.7 – Frontend Apache Tomcat ver. 7.0 – Server

3. Apache Tomcat ver. 7.0 – Server

4. Navicat 8.0 for MySQL – Database

3.3 Environmental Setup:

The basic setup of our system includes clients and server. We have successfully implemented this paper on experimental basis using VMware and LAN network.
1. A Virtual Cloud Environment in set up using one server computer and many client computers in LAN network in the form wireless connection using APACHE TOMCAT.

2. Can also be run on single system using virtual network like VMware

3.4 Algorithm:

In order to maintain the security of the file, we have used AES algorithm (Advance Encryption Standard) which uses a key for encryption.Also we are using **Proof of Ownership** which is a protocol used by a server to verify that a file is owned by a particular client.

3.5 Architecture:

We are following **MVC (Model** View Controller) architecture in our system in which **JAVA** is used as model, **JSP** (Java Server Pages) is the View and Controller is **Servlet.**

All the coding and data is managed by java, user enters data in JSP pages on the browser and the data is then sent to servlet for further processing. The servlet data is transferred to **DAO** (Data Access Object) layer.

The **DAO** layer fires queries related to data as per the controllers request (servlet).The data is either stored or fetched from the database as per the queries fired. The response is sent to servlet. The servlet (controller) then forwards the response or redirects to the required page.

We have used Bootstrap to create dynamic web pages based on xml, html etc. User's request is then sent to the Server. Server is a controller that decides whether to process forward the request or redirect back. Any request raised is processed first by the servlet. If the request response is yes it is then sent to **DAO** (Direct Access Object) which is connected to the database**. DAO** is an object that imparts abstract interface to various types of database. Once mapping on application calls is done, DAO gives specific data operations without revealing details of database.

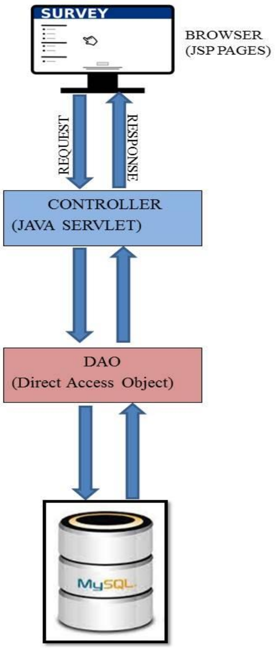


Fig – 3: Request-Response Flow

4. Working Module

4.1 Registration:

We use a hybrid cloud in which we store data of various department of an organization. When a new user wants to log in to access the cloud he initially has to register himself to a particular department in the cloud using his credentials. Even after registration a user cannot login until he is authorized to be 'active' by other existing authorized user of thatdepartment. Once the authorized user grants him the permission the new user turns authorized and can access cloud.

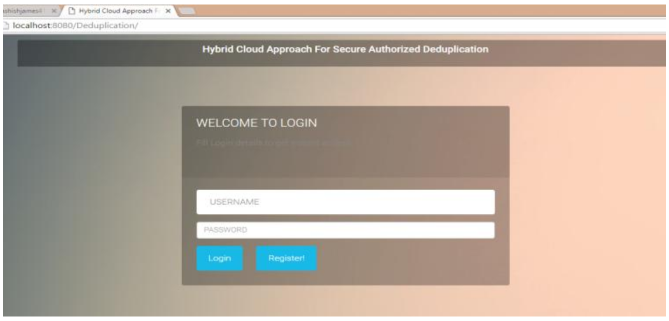


Fig – 4: Login Page

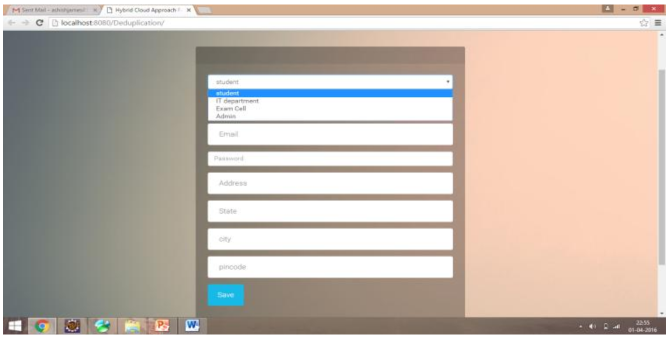


Fig - 5: Registration Page

4.2 Uploading a file:

Only a registered user can upload a file. While uploading a file the user can browse and select the file to be uploaded. Once the file is selected the user can select the departments to which the file should be visible in short the user can customize the visibility of the file. No other departments accept the selected ones can view the file.

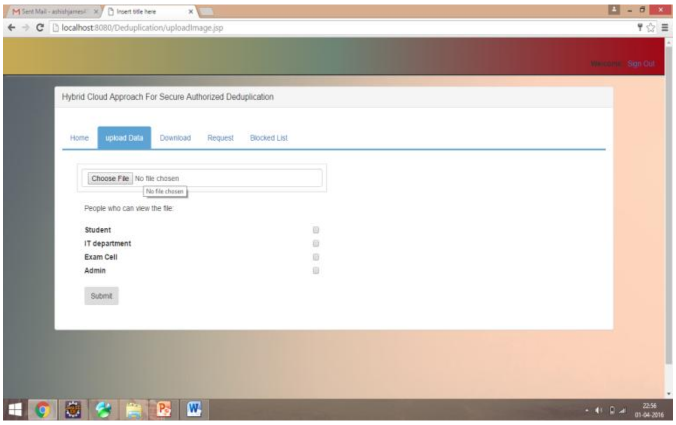
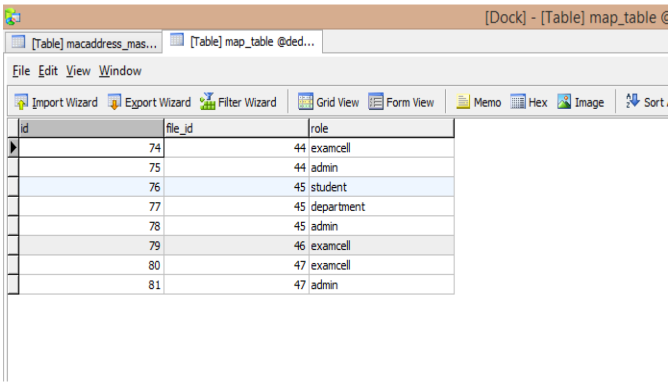

localhost:8080/Deduplication/uploadImage.jsp
Hybrid Cloud Approach For Secure Authorized Deduplication
Home
upload Data
Download
Request
Blocked List
Choose File
No file chosen
People who can view the file:
Student
IT department
Exam Cell
Admin
Submit

Fig – 6: File Uploading by providing Access Rights

4.3 Deduplication:

When a user uploads a file, in the backend the file gets split into blocks consisting of certain no. of lines of the file. Each block is then read and encrypted using AES algorithm. The encrypted block is compared to all the other blocks in the cloud to avoid repetition of data on cloud. Each file is given a unique ID and each block of the file is stored in an array with File ID as its name. If some block tends to be repeated the address of the existing block is pointed and stored in the array.



id	file_id	role
74	44	examcell
75	44	admin
76	45	student
77	45	department
78	45	admin
79	46	examcell
80	47	examcell
81	47	admin

Fig – 7: Deduplication of the files with same content (Backend)

4.4 Downloading a file:

While downloading some file from the cloud the is first asked for OTP (One Time Password) which he receives on his registered e-mail ID .After he has entered the correct OTP he can view and download only those file made visible to department.

When a user wants to download the file at the backend the file is searched using unique ID given to it. In the array that contains all the encrypted blocks and pointers of the file are called, each block is then decrypted, read and finally merged together.

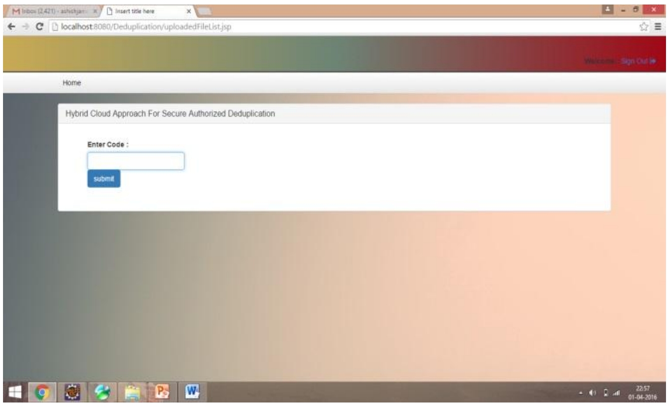
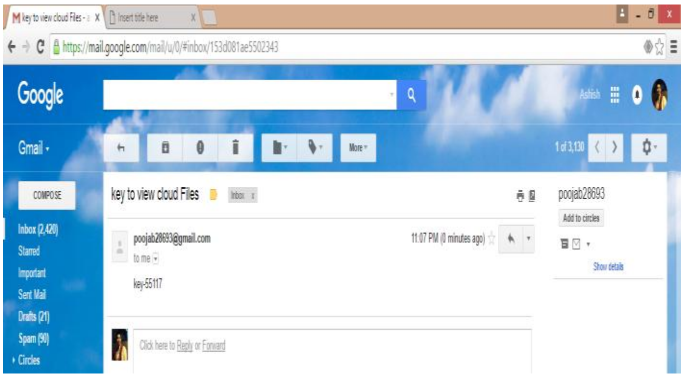

localhost:8080/Deduplication/uploadedFileList.jsp
Home
Hybrid Cloud Approach For Secure Authorized Deduplication
Enter Code :
submit

Fig – 8: Demand for OTP for downloading File



https://mail.google.com/mail/u/0/#inbox/153d081ae5502343

key to view cloud Files

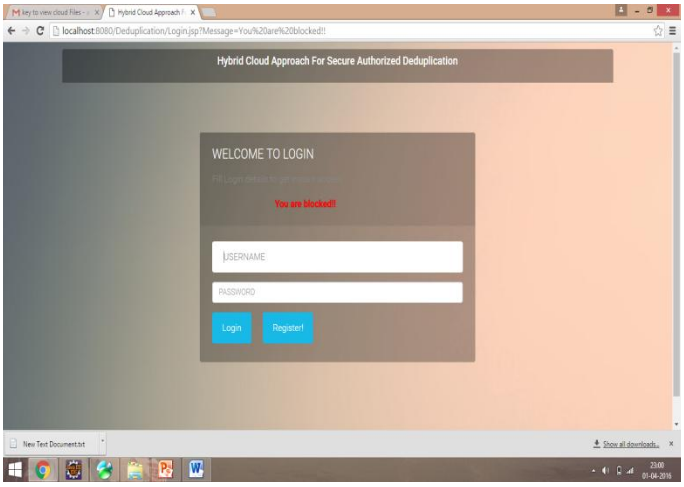
poojab28693@gmail.com
to me
11:07 PM (0 minutes ago)

key-55117

Fig – 9:OTP received via mail

4.5 Security:

Suppose a hacker tries to distort or hack some file from the cloud as an authorized user, he will be asked for **OTP** which is received on the mail Id via a message on the phone of the registered person which the hacker is unaware of .The hacker would try and make random guesses of the **OTP**. After his three wrong attempts the user is blocked and becomes unauthorized to login into the cloud. Also the **MAC** address and the **IP** address of the machine from which the activity took place is traced and mailed to the user. The user remains unauthorized and inactive till another authorized user of that department grants him permission to become active.



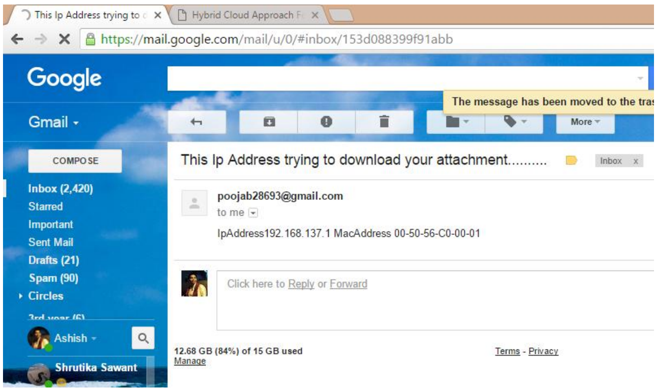
localhost:8080/Deduplication/Login.jsp?Message=You%20are%20blocked!!

Hybrid Cloud Approach For Secure Authorized Deduplication

WELCOME TO LOGIN

You are blocked!!

Fig – 10: User being block after 3 wrong attempts



https://mail.google.com/mail/u/0/#inbox/153d088399f91abb

The message has been moved to the tras

This Ip Address trying to download your attachment..........

poojab28693@gmail.com
to me

IpAddress192.168.137.1 MacAddress 00-50-56-C0-00-01

12.68 GB (84%) of 15 GB used

Fig – 11: IP and **MAC** address of the hacker,s machine sent of registered users email.

5. CONCLUSION

Hence, with the help of this paper we can solve various problems like managing storage of increasing data in the cloud, security and confidentiality of the files stored in cloud and also introducing features like user friendly and easy to understand. Since we have successfully implemented this paper on virtual environment, we would now implement it in a areal time scenario like large industries and institutions. We would also make advancements to it on real world scenario like figure prints for verification, providing OTPs on text message and many more.

REFERENCES

[1] Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou, ”A HYBRID CLOUD APPROACH FOR SECUR AUTHORIZED DEDUPLICATION”M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

[2] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg. Proofs of ownership in remote storage systems. In Y. Chen, G. Danezis, and V. Shmatikov, editors, ACM Conference on Computer and Communications Security, pages 491–500. ACM, 2011..

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

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

[5] M. Bellare, S. Keelveedhi, and T. Ristenpsart. Message-locked encryption Band secured duplication. InEUROCRYPT,pages 296 – 312, 2013.

[6] M. Bellare, C. Namprempre, and G. Neven. Security proofs for identity-based identification and signature schemes. J. Cryptology, 22(1):1–61, 2009.

[7] 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.

BIOGRAPHIES

Taranpreet Bhatti
"Currently perusing BE Engineering Degree in Information Technology at K.C College of Engineeringand Management Studies & Research affiliated to Mumbai University"

Ashish James
"Currently perusing BE Engineering Degree in Information Technology at K.C College of Engineering and Management Studies & Research affiliated to Mumbai University"

Siddhi Narvekar
"Currently perusing BE Engineering Degree in Information Technology at K.C College of Engineering and Management Studies & Research affiliated to Mumbai University"