SECURE AND EFFICIENT DATA DEDUPLICATION IN JOINT CLOUD STORAGE

A. Durga Devi [1], Kopparthi Sai Teja[2],

[1]**Assistant professor , PG DEPT,** Dantuluri Narayana Raju College**, Bhimavaram, Andharapradesh**
Email:- adurgadevi760@gmail.com
[2]**PG Student of MCA, Dantuluri** Narayana Raju College**, Bhimavaram, Andharapradesh**
Email:-saitejak202@gmail.com

ABSTRACT

Data deduplication can efficiently eliminate data redundancies in cloud storage and reduce the bandwidth requirement of users. However, most previous schemes depending on the help of a trusted key server (KS) are vulnerable and limited because they suffer from revealing information, poor resistance to attacks, great computational overhead, etc. In particular, if the trusted KS fails, the whole system stops working, i.e., single-point-of-failure. In this paper, we propose a Secure and Efficient data Deduplication scheme (named SED) in a Joint Cloud storage system which provides the global services via collaboration with various clouds. SED also supports dynamic data update and sharing without the help of the trusted KS. Moreover, SED can overcome the single-point-of- failure that commonly occurs in the classic cloud storage system. According to the theoretical analyses, our SED ensures the semantic security in the random oracle model and has strong anti- attack ability such as the brute-force attack resistance and the collusion attack resistance. Besides, SED can effectively eliminate data redundancies with low computational complexity and communication and storage overhead. The efficiency and functionality of SED improves the usability in client-side. Finally, the comparing results show that the performance of our scheme is superior to that of the existing schemes.

1 INTRODUCTION

CLOUD storage is a platform to provide large scale data storage and service access at a ―pay-as- you-go‖ fashion. However, a lot of redundant data in cloud storage has seriously wasted and occupied storage resources. Data deduplication is an effective technology to detect and eliminate redundant data [1]. After that, only a single copy of the data is uploaded and stored. Thus, the data

/deduplication technology can reduce the bandwidth requirement of client-side and improve the space utilization efficiency of server-side. Currently, it has widely used in various cloud computing services to improve user experience and save storage space.

The classic data deduplication scheme and its variants [2], [3], [4], [5], [6], [7], where the framework consists of a key server (KS), a cloud storage provider (CSPs), and users, ensure the security depending on the trusted KS. What is worse, these classic schemes may suffer from the single-point-of-failure and ―platform lock-in‖ issues. If the trusted KS fails, the cloud storage system stops working and data outsourcing protocols cannot be

implemented. Recently, a new model of cloud computing, called as Joint- Cloud computing system [8], has been designed to solve the above-mentioned issues well. The network architecture of JointCloud consists of users and multiple CSPs providing various services.

Literature Survey

OUTLINE FOR LITERATURE SURVEY:

Introduction

- **Definition and Importance**: Explain what data deduplication is and why it is important for data storage and management.
- **Scope and Objectives**: Outline the scope of your survey and what you aim to achieve.

Background

Data Deduplication Techniques: Overview of different data deduplication methods (e.g., file-level, block-level, byte-level).

Security Concerns: Discuss the security challenges associated with data deduplication, including privacy and data integrity issues.

3 IMPLEMENTATION STUDY

EXISTING SYSTEM:

Convergent encryption is one of the main approaches to ensure the security of data in deduplication, which can protect the outsourced data against the untrusted andalicious CSPs. Bellare et al: [10] formalized a primitive as message-locked encryption (MLE) scheme. Then, some variants [12], [13] were proposed based on the work of Bellare. However, these MLE-based schemes were facing many potential risks because the keys used to encrypt files are derived from the files themselves. Abadi et al: [11] designed a full randomized scheme and a deterministic encrypted scheme for bounded message distributions based on the non-degenerate efficiently computable bilinear map. Li et al: [15] presented a scheme to achieve reliable key management in deduplication. Then, Jiang et al: [14] showed a secure scheme for deduplication with the randomized tag.

Disadvantages:

- There is no Intra-deduplication which just considers that the data owner has outsourced his/her data by the same KS. It is more efficient for the backup system which is not in an existing system.

Proposed System & alogirtham

In this paper, we propose a secure and efficient data deduplication scheme SED without the help of the trusted KS in the JointCloud storage system. Some sub-algorithms of our SED are inspired by the fully randomized tag generation algorithm [11] which helps with duplicates detection and protects the outsourced data against the collusion attacks. Different from the previous deduplication schemes, our SED ensures that the ciphertext and the tag can satisfy semantic security. Any adversary cannot get any useful information from the tag and ciphertext. Moreover, our SED is the first scheme that supports data update and data sharing securely.

4.1 Advantages:

- The encryption algorithm and the tag generation algorithm of the proposed SED ensure the semantic security. Moreover, SED can resist the typical attacks such as the brute-force attack, tampering attack, and collusion attack.
- The SED implements secure deduplication without the help of the trusted key server. It also supports data updating and sharing cross-clouds. Furthermore, SED solves the single-point-of- failure issue and improves the scalability of the classic deduplication scheme.

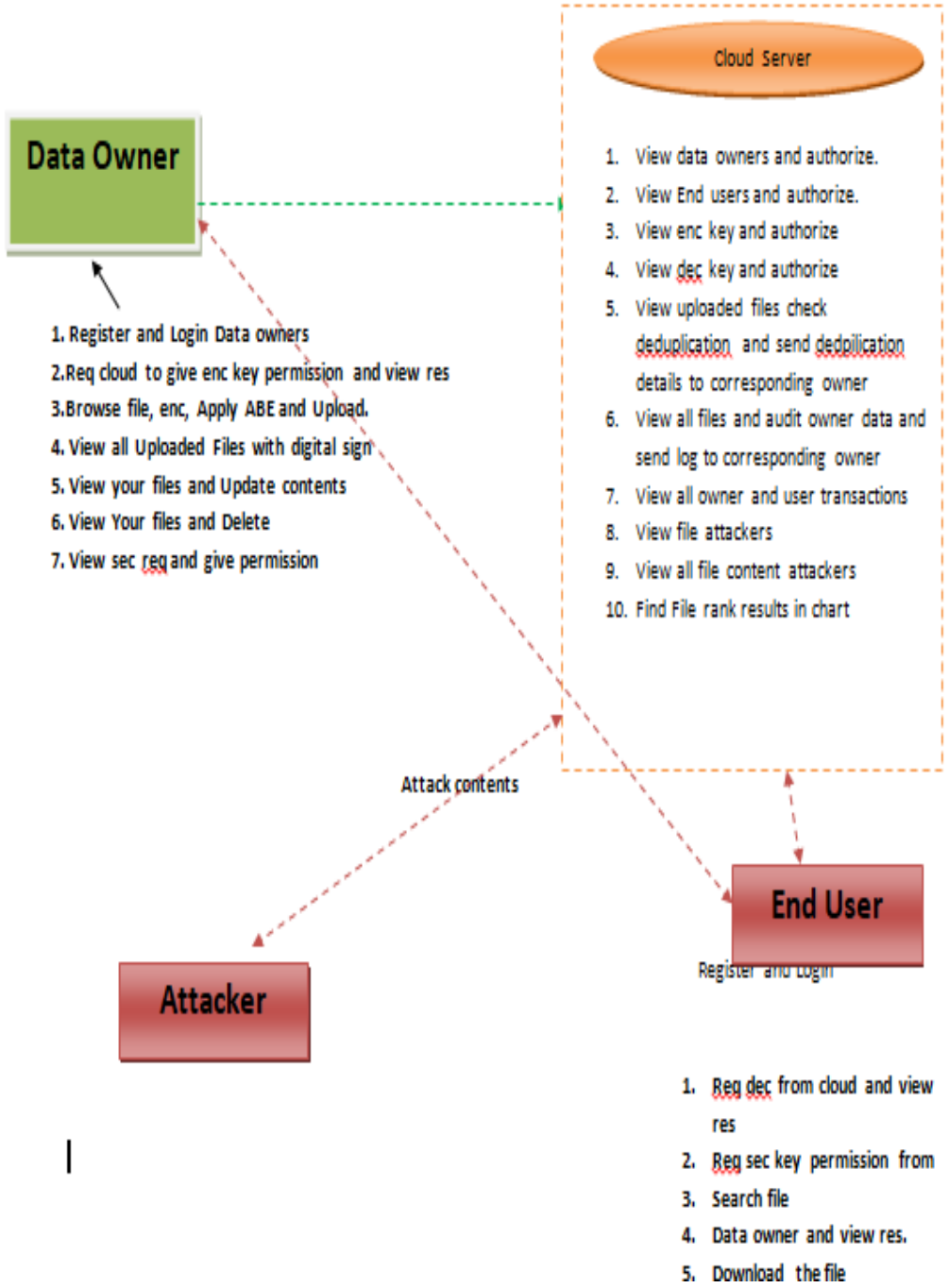


Fig:3.1 System Architecture

IMPLEMENTATION

- **Data Owner**

  In this module, the data owner uploads their data in the cloud server. For the security purpose the data owner encrypts the file and the index name and then store in the cloud. The data encryptor can have capable deleting of a specific file. And also he can view the transactions based on the files he uploaded to cloud.

- **Data User**

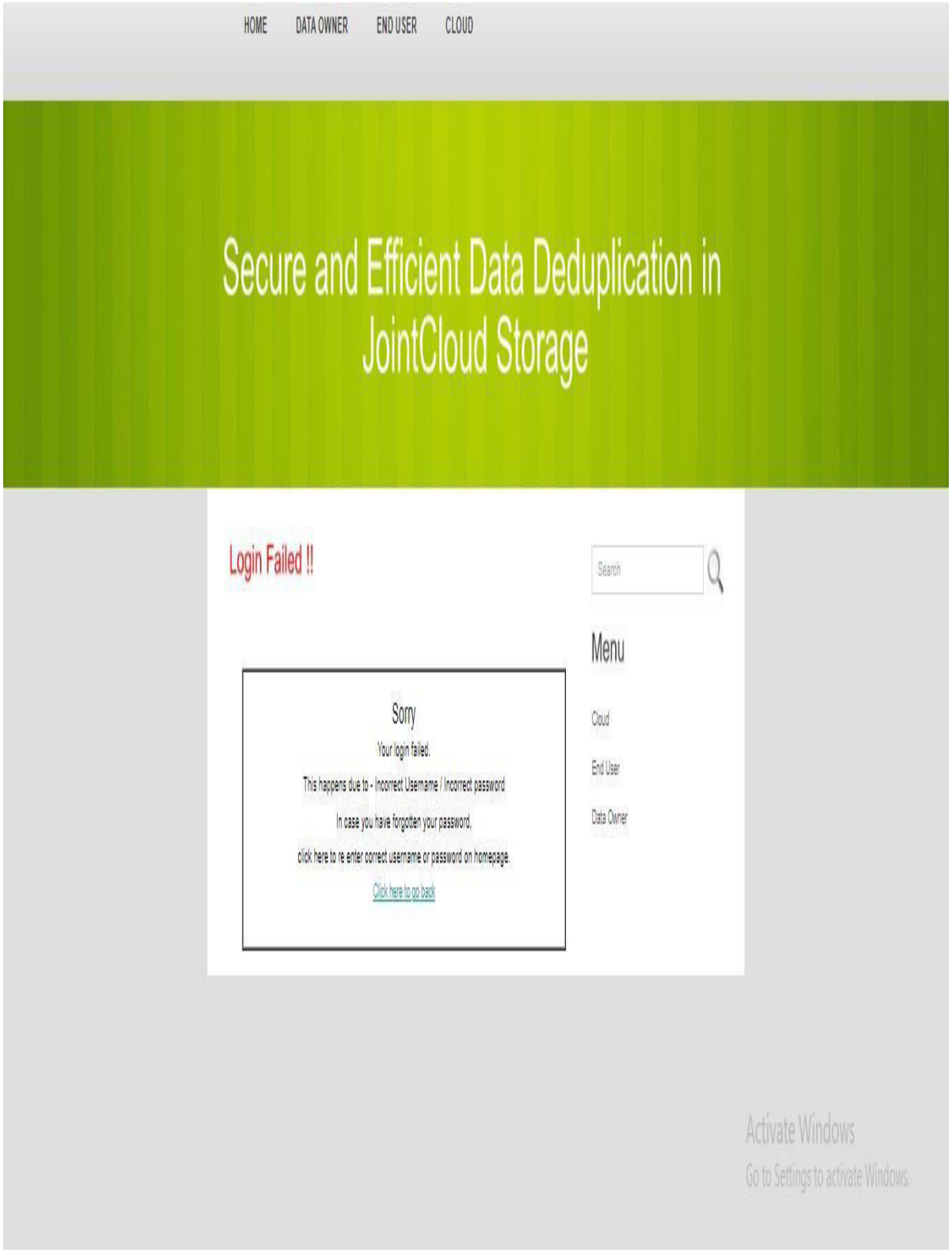
  In this module, user logs in by using his/her user name and password. After Login user requests search control to cloud and will Search for files based on the index keyword with the Score of the searched file and downloads the file. User can view the search ratio of the files and also the top k documents.

5 RESULTS AND DISCUSSION

SCREEN SHORTS:

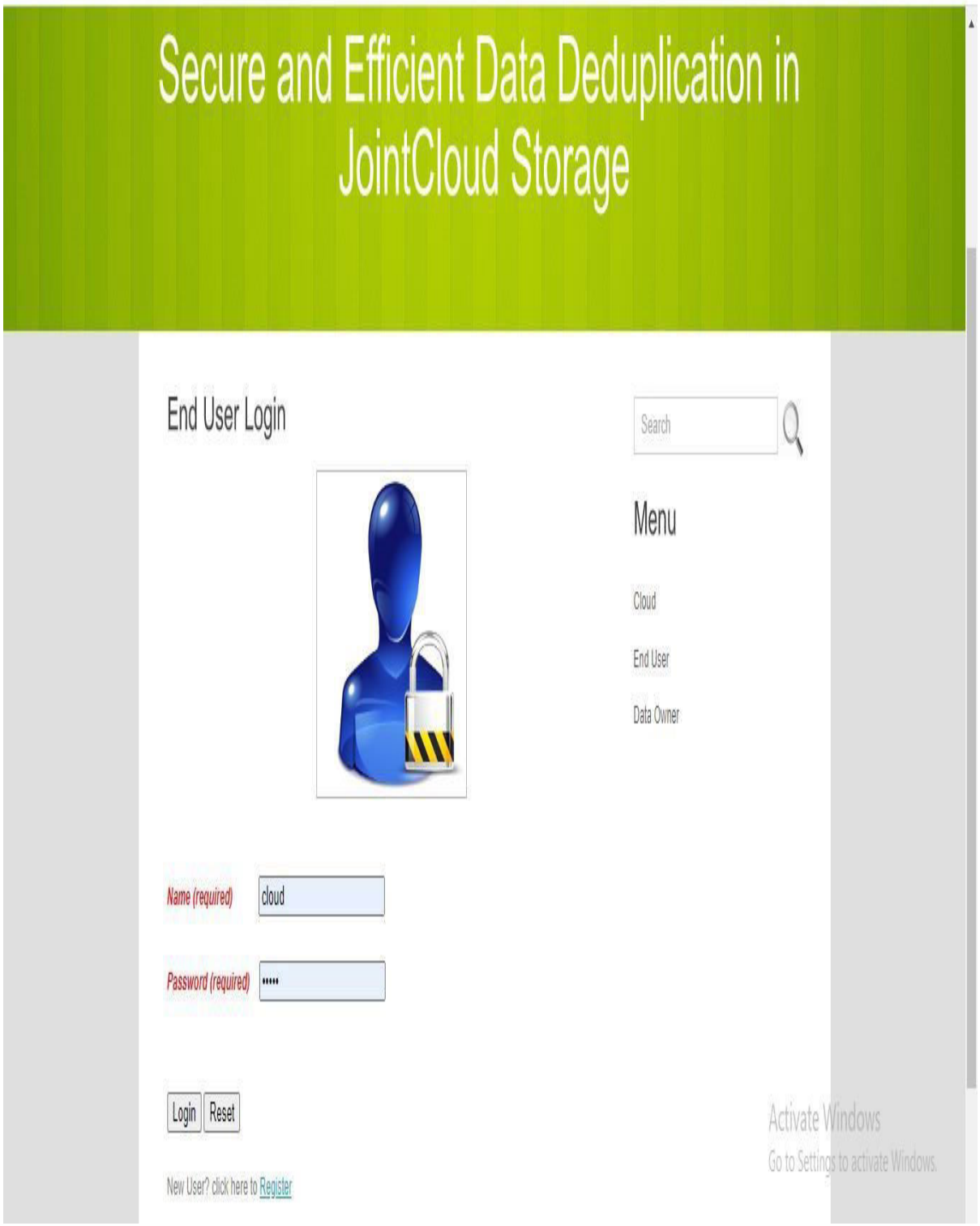
HOME PAGE:

WRONG LOGIN:

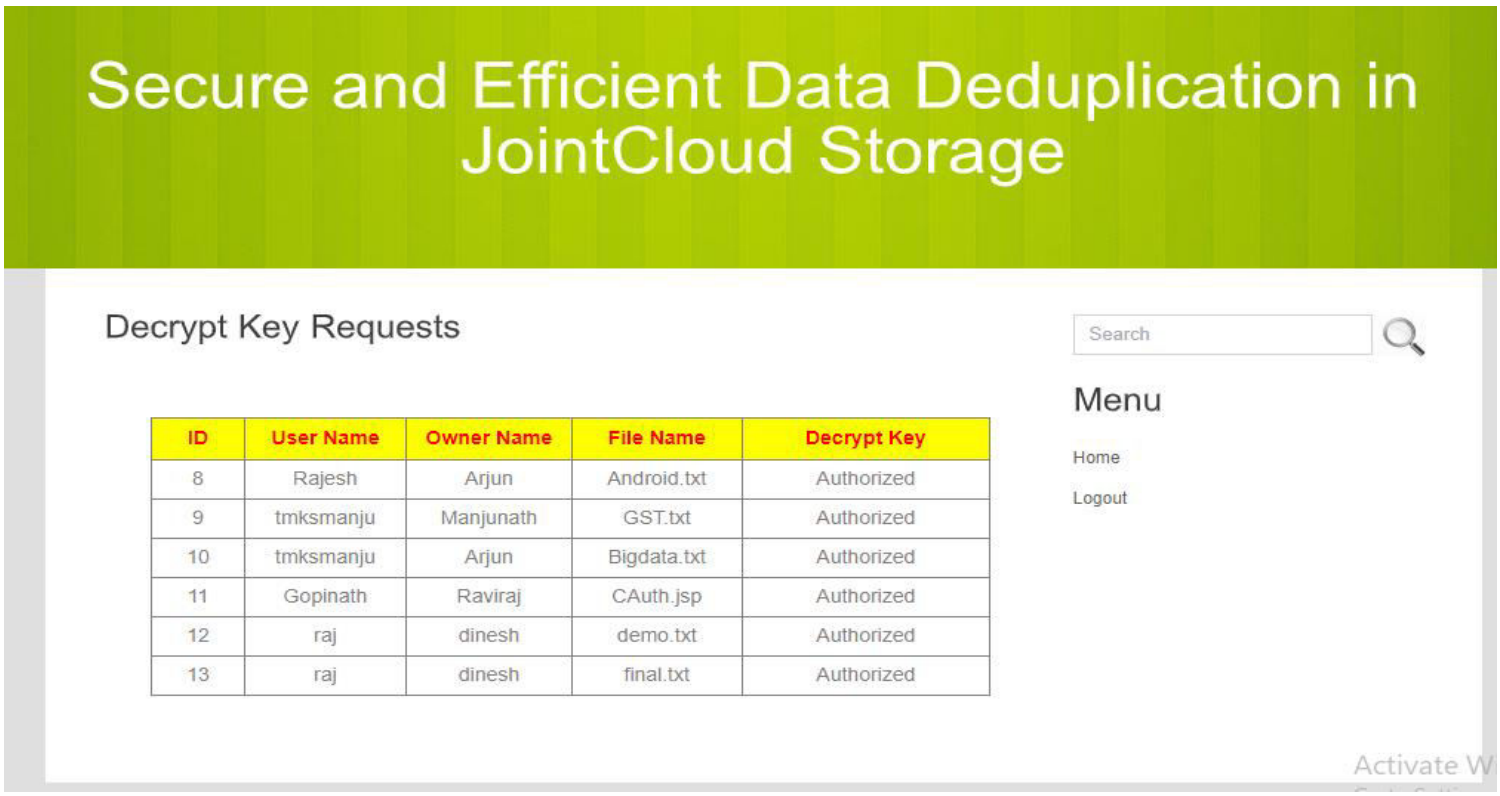

HOME
DATA OWNER
END USER
CLOUD
Secure and Efficient Data Deduplication in JointCloud Storage
Login Failed !!
Search
Menu
Cloud
End User
Data Owner
Sorry
Your login failed.
This happens due to - Incorrect Username / Incorrect password
In case you have forgotten your password,
click here to re enter correct username or password on homepage.
Click here to go back
Activate Windows
Go to Settings to activate Windows.

DATA OWNER LOGIN:

END USER LOGIN:

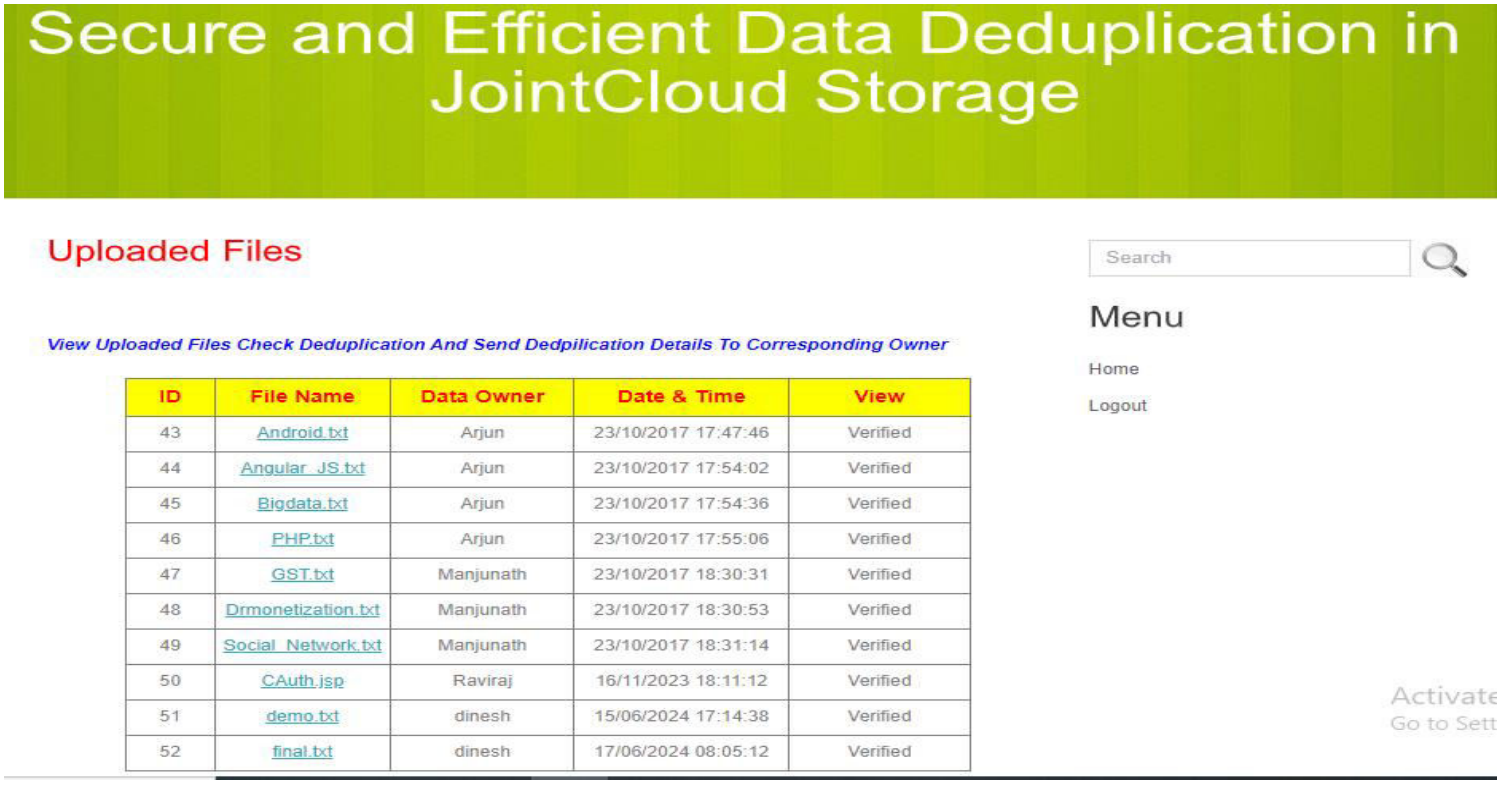


DECRIPT KEY REQUESTS:



ID	User Name	Owner Name	File Name	Decrypt Key
8	Rajesh	Arjun	Android.txt	Authorized
9	tmksmanju	Manjunath	GST.txt	Authorized
10	tmksmanju	Arjun	Bigdata.txt	Authorized
11	Gopinath	Raviraj	CAuth.jsp	Authorized
12	raj	dinesh	demo.txt	Authorized
13	raj	dinesh	final.txt	Authorized

UPLOAD FILES:



ID	File Name	Data Owner	Date & Time	View
43	Android.txt	Arjun	23/10/2017 17:47:46	Verified
44	Angular_JS.txt	Arjun	23/10/2017 17:54:02	Verified
45	Bigdata.txt	Arjun	23/10/2017 17:54:36	Verified
46	PHP.txt	Arjun	23/10/2017 17:55:06	Verified
47	GST.txt	Manjunath	23/10/2017 18:30:31	Verified
48	Drmonetization.txt	Manjunath	23/10/2017 18:30:53	Verified
49	Social_Network.txt	Manjunath	23/10/2017 18:31:14	Verified
50	CAuth.jsp	Raviraj	16/11/2023 18:11:12	Verified
51	demo.txt	dinesh	15/06/2024 17:14:38	Verified
52	final.txt	dinesh	17/06/2024 08:05:12	Verified

CLOUD LOGIN:

FILE ATTACKERS:

6. CONCLUSION AND FUTURE WORK

CONCLUSION

In this paper, we have designed a secure and efficient scheme SED for data deduplication without the help of the trusted KS. The proposed SED has reduced the communication and computation overhead of client-side and improved efficiency based on the CDH problem in the JointCloud storage system. Its concise algorithms of encrypting and generating tag satisfy the semantic security and the tag consistency (including security and validity), respectively. Moreover, SED improves the scalability and solves the single-point-of-failure of KS in the classic cloud storage system. SED has strong capacity against typical attacks such as the brute-force attack and the collusion between malicious CSPs and unauthorized users. Besides, SED supports dynamic supports data operations, including deletion, modification, and sharing, which improves the functionality and usability. To the best of our knowledge, SED is the first scheme considering the case that data owner shares his/her outsourced data to the permitted users. According to the theoretical and experimental analyses, our SED is secure and has low computation, communication, and storage complexity. From the comparison with the previous scheme, our SED is more secure, efficient, and functional.

7. REFRENCES

1] P. Christen, ―A survey of indexing techniques for scalable record linkage and deduplication,‖ IEEE Transactions on Knowledge and Data Engineering, vol. 24, no. 9, pp. 1537–1555, 2012. [2] G. Jia, G. Han, J. J. P. C. Rodrigues, J. Lloret, andW. Li, ―Coordinate memory deduplication and partition for improving performance in cloud computing,‖ IEEE Transactions on Cloud Computing, vol. 7, no. 2, pp. 357–368, 2019.[3] W. Xia, X. Zou, H. Jiang, Y. Zhou, C. Liu, D. Feng, Y. Hua, Y. Hu,and Y. Zhang, ―The design of fast content- defined chunking for data deduplication based storage systems,‖ IEEE Transactions on Parallel and Distributed Systems, vol. 31, no. 9, pp. 2017–2031, 2020.[4] J. Li, J. Li, D. Xie, and Z. Cai, ―Secure auditing and deduplicatingdata in cloud,‖ IEEE Transactions on Computers, vol. 65, no. 8, pp. 2386–2396, 2016. [5] L. Liu, Y. Zhang, and X. Li, ―Keyd: Secure key-deduplication withidentity-based broadcast encryption,‖ IEEE Transactions on Cloud Computing, pp. 1–1, 2018. [6] J. Ni, K. Zhang, Y. Yu, X. Lin, and X. S. Shen,

―Providing task allocation and secure deduplication for mobile crowdsensing via fog computing,‖ IEEE Transactions on Dependable and Secure Computing, pp. 1–1, 2018.

[7] Y. Zheng, X. Yuan, X.Wang, J. Jiang, C.Wang, and X. Gui, ―Towar encrypted cloud media center with secure deduplication,‖ IEEE Transactions on Multimedia, vol. 19, no. 2, pp. 251–265, 2017. [8] H. Wang, P. Shi, and Y. Zhang, ―Jointcloud: A cross-cloud cooperation architecture for integrated internet service customization,‖ in 2017 IEEE 37th International Conference on Distributed Computing Systems, 2017, pp. 1846–1855. [9] K. Huang, X. Zhang, Y. Mu, F. Rezaeibagha, X. Wang, J. Li, Q. Xia, and J. Qin, ―Eva: Efficient versatile auditing scheme for iot-based datamarket in jointcloud,‖ IEEE Internet of Things Journal, vol. 7, no. 2, pp. 882–892, 2020. [10] M. Bellare, S. Keelveedhi, and T. Ristenpart, ―Message- locked encryption and secure deduplication,‖ in International Conference on the Theory and Applications of Cryptographic Techniques, 2013, pp. 296–312. [11] M. Abadi, D.Boneh, I. Mironov, A. Raghunathan, and G. Segev,―Message-locked encryption for lock-dependent messages,‖ in Advancesin Cryptology – CRYPTO 2013. Springer Berlin

Heidelberg, 2013,
pp. 374–391. [12] M. Bellare and S. Keelveedhi, ―Interactive message-locked encryption and secure deduplication,‖ in Public-Key Cryptography – PKC 2015, J. Katz, Ed. Springer Berlin Heidelberg, 2015, pp. 516–538. [13] M. Bellare, S. Keelveedhi, and T. Ristenpart, ―Dupless: serveraided encryption for deduplicated storage,‖ in Usenix Conference on Security, 2013, pp. 179–194.