International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

Privacy Protection Using Deduplication in Cloud

Dr.Subba Rao Kolavennu
(Assistant Professor, Department of Computer Science and Engineering, Sphoorthy Engineering College, Hyderabad.)
Email: profrao99@gmail.com
M.Manisha, B.Tech
(Student, Department of Computer Science and Engineering, Sphoorthy Engineering College, Hyderabad.)
Email: 19n81a0570manishayadav@gmail.com
G.Ramya, B.Tech
(Student, Department of Computer Science and Engineering, Sphoorthy Engineering College, Hyderabad.)
Email: 19n81a0575ramya@gmail.com
K.V.C.Sarika, B.Tech
(Student, Department of Computer Science and Engineering, Sphoorthy Engineering College, Hyderabad.)
Email: 19n81a0578sarika@gmail.com
V.Teja Sri, B.Tech
(Student, Department of Computer Science and Engineering, Sphoorthy Engineering College, Hyderabad.)
Email: 19n81a0585teju@gmail.com

ABSTRACT:

Cloud storage auditing with deduplication is a crucial technique for ensuring data integrity while minimizing storage overhead by storing only one instance of duplicated files. However, existing auditing schemes with deduplication are susceptible to brute-force dictionary attacks, compromising user privacy. This paper introduces cloud storage auditing scheme with deduplication that provides robust privacy protection. The proposed scheme safeguards user file privacy, even for predictable or small-space files, by employing innovative methods for file index generation and encryption key derivation. The scheme offers lightweight computation for generating data authenticators, verifying data integrity, and enabling seamless file retrieval. The security analysis and performance evaluation demonstrate that the proposed scheme achieves desirable security and efficiency, addressing the limitations of existing approaches.

Keyword: Cloud storage auditing, deduplication, privacy protection, file index generation, encryption key derivation, lightweight computation, data authenticators, security analysis, performance evaluation.

1.INTRODUCTION:

With the explosive growth of digital data, deduplication techniques are widely employed to backup data and minimize network and storage overhead by detecting and eliminating redundancy among data. Instead of keeping multiple data copies with the same content, deduplication eliminates redundant data by keeping only one physical copy and referring other redundant data to that copy. Deduplication has received much attention from both academia and industry because it can greatly improves storage utilization and save storage space, especially for the applications with high deduplication ratio such as archival storage systems.Though deduplication technique can save the storage space for the cloud storage service

providers, it reduces the reliability of the system. Data reliability is actually a very critical issue in a deduplication storage system because there is only one copy for each file stored in the server shared by all the owners.Encryption mechanisms have usually been utilized to protect the confidentiality before outsourcing data into cloud. Most commercial storage service provider is reluctant to apply encryption over the data because it makes deduplication impossible. The reason is that the traditional encryption

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

mechanisms, including public key encryption and symmetric key encryption, require different users to encrypt their data with their own keys. As a result, identical data copies of different users will lead to different ciphertexts. To solve the problems of confidentiality and

deduplication, the notion of convergent encryption has been proposed and widely adopted to enforce data confidentiality while realizing deduplication.

2.LITERATURE SURVEY:

We have surveyed the existing projects and finally thought of making necessary modifications for getting the latest edition

EXISTING SYSTEM:

Users usually encrypt their data before outsourcing them to the cloud since they would not like to disclose their sensitive data to the cloud and other parties In order to realize deduplication over encrypted data, the convergent encryption (CE) was proposed to encrypt data. A convergent encryption algorithm encrypts data with a key deterministically derived from the data (e.g., the file's hash value). Thus, the same file will produce the same ciphertext. It means that the deduplication over ciphertexts is feasible.

In addition, all users who want to upload file to the cloud need to generate a file index and send it to the cloud for duplicate check. With the file index, the cloud can verify whether the file uploaded by the user is duplicated or not. If the file index has been kept by the cloud, then the subsequent users do not need to upload data to the cloud any more. Most of deduplication schemes set the hash value of the file as the file index. It will result in the data privacy leakage because the malicious cloud or other parties might guess or derive the content of file by performing the brute-force dictionary attacks.Thus, how to realize deduplication supporting strong privacy protection in cloud storage auditing is very important and valuable. Unfortunately, previous schemes are weak in privacy protection because they cannot fully defend against the brute-force dictionary attacks.

Disadvantages of Existing System:

Only one copy for each file stored in cloud even if such a file is owned by a huge number of users and There are brute force attacks and chance of leakage of user privacy.

PROPOSED SYSTEM:

In this paper, we investigate how to fully resist the bruteforce dictionary attacks and realize deduplication with strong privacy protection in cloud storage auditing, and propose a concrete scheme satisfying this property. In order to realize deduplication with strong privacy protection, we design a novel method to generate the file index, and employ a new strategy to generate the key for file encryption. In the detailed design, the file index is generated with the help of an Agency Server (AS) instead of directly being produced by the hash value of file.

The key for file encryption is generated with the file and the file label. The file label is kept by the user secretly. In this way, the privacy of the user's file is protected against the cloud and the AS. In order to improve the storage efficiency, the users, who own the same file, are able to generate the same ciphertext and the same authenticators. The proposed scheme effectively achieves data deduplication and authenticator deduplication. Furthermore, to reduce the computation burden on the user side, the user only needs to perform lightweight computation to generate data authenticators, verify the integrity of the cloud data, and retrieve his file from the cloud. We give the security analysis of the proposed scheme, showing that the proposed scheme satisfies correctness, soundness and strong privacy protection. We also justify the performance by concrete implementations. The result shows that the proposed scheme is efficient.

Advantages of Proposed System:

Data deduplication is a process that eliminates excessive copies of data and significantly decreases storage capacity requirements and Users data is secure and no brute force attacks.

3.IMPLEMENTATION:

We have implemented two storage system prototypes to compare the performance overhead of our proposition with

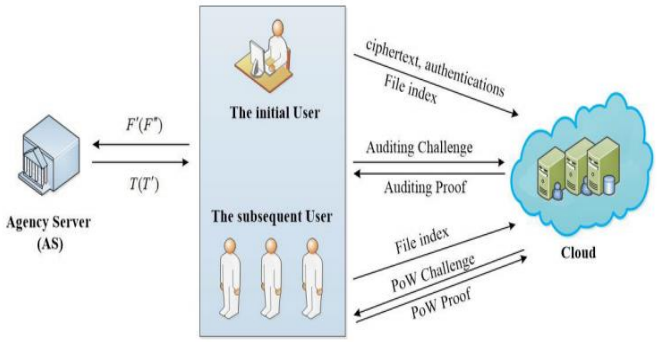
International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

respect to a classic storage system with no data encryption. Specifically, we have developed a classic storage system with a client and storage server software modules, and one implementing our proposition with a client, a deduplication proxy and storage server software modules.

All these software modules are implemented in python 2.7.6 and access the pycrypto library for the cryptographic operations. We use the SHA256 algorithm as the hash function, RSA1024 for asymmetric encryption operations, and AES for the symmetric encryption operations with keys that are 256 bits long. The storage servers use the MongoDB3 database to store the meta-data of the stored files as well as files owners. The software modules are executed on three different virtual machines (VMs) running on Ubuntu 12.04.4 LTS with 1GB of memory and an AMD Opteron(TM) octa-core Processor 6220@3GHz. The network topology is as follows: the VM executing the DP software is located on the network path between the VM running the different clients modules and the one executing the different SSs modules.

The average duration values of the these operations. We observe that the intra-user deduplication in our scheme consumes more communication resources than an inter-user deduplication in a classic scheme due to the use of a DP (around 18ms of overhead for a file of 64MB length). However most of the overhead comes from the encryption key creation, the encryption key encryption4 and the file encryption which depends on the file size. We also observe that operations involving the interuser deduplication in our scheme have higher durations than the ones involving an intra-user deduplication. This is due to (i) the overhead of communications when applying the interuser deduplication because the file has to be uploaded to the DP and then a reference is uploaded to the the SS and (ii) the added delay by the DP to make unnoticeable the inter-user deduplication to client. Actually in our scheme, a put operation involving an inter-user deduplication has a similar duration than a file uploads from the client to the SS.

4.APPLICATION ARCHITECTURE:



5.RESULTS AND ANALYSIS:

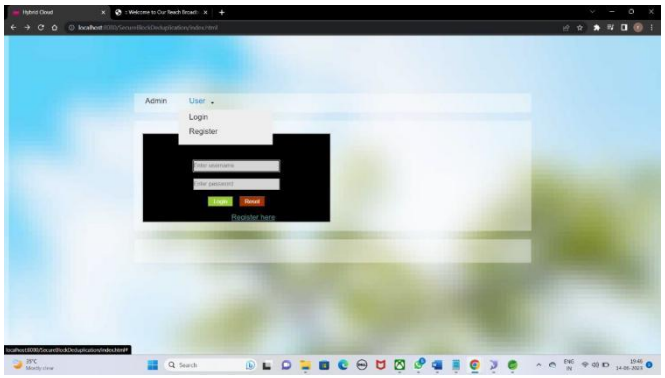
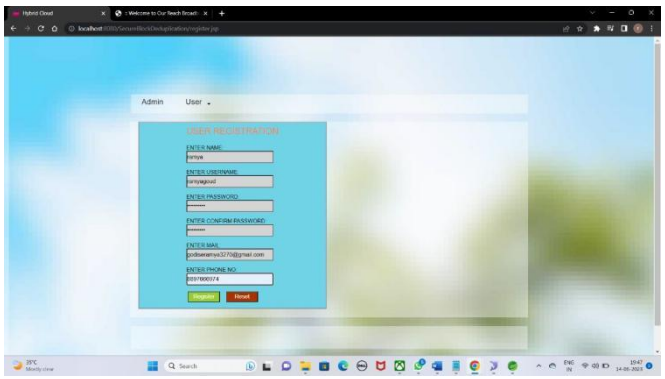
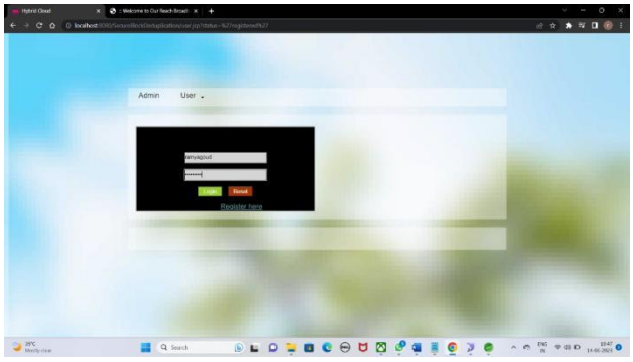
The proposed scheme successfully overcomes brute fore attacks and slove user privacy.Overall, the results demonstrate that the proposed scheme effectively addresses the problem of user privacy leakage in cloud storage auditing with deduplication, offering improved storage efficiency and computational efficiency compared to existing schemes.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

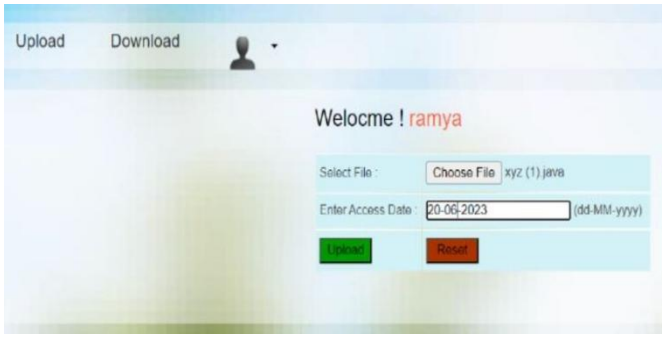
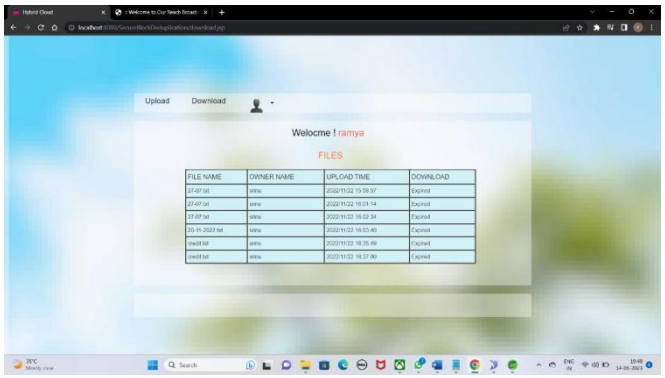
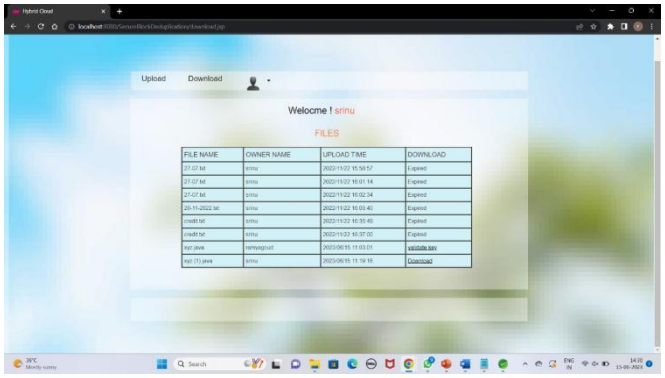
Test ID	Test Name	Inputs	Process	Excepted Output	Actual Output	Status
1	Login Test	UserName, Password	Validate Username and password on database	Need to redirect to user home page	It's Redirected to user home	Success
2	Registration Test	Username, password, Mobile number, email etc…	insert the users into database	Need to insert the user details into database	It's inserted	Success
3	Upload file	File,UserId.	Insert the files into the cloud	Need to insert the files into the database	It's inserted	Success
4	Grant access	User id, Permission	Grant the access to the user to access the files	Need to grant permission to user	permission granted	Success
5	Send Mail	User id, mailed, key, message	Send these inputs to specified user mail id	Need to send the details to mail id	Mail forwarded	success
6	Download File	Userid, file id,key,permission	If given user having access to files and if he entered the	Download the file if he specified the right details	File downloaded	Success

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

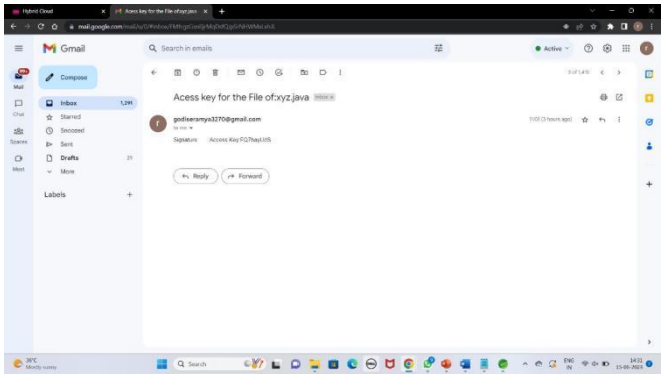
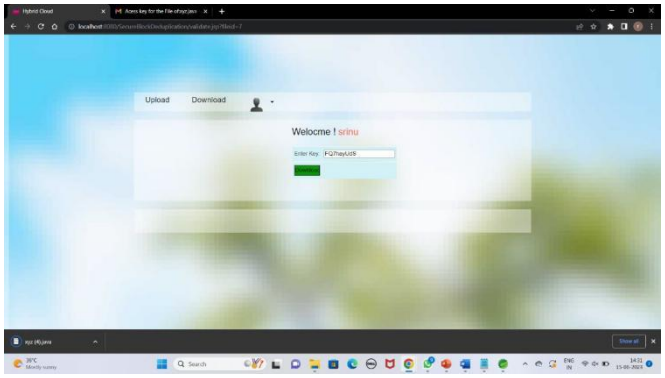
			Wright file key nee do download the file			

UI DESIGN	DESIGN DESCRIPTION
	The above screen appears after clicking http://localhost:8080/SecureBlockDeduplication/index.htmlcloud .
	The screen shows the registration page. The user need to enter the required details shown.
	After registration the user need to login the application by giving username, password.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

	The screen shows upload page.The user can upload files by clicking on the upload.
	The screen shows download page .The user can see all the uploaded files .
	The screen shows all the user uploaded files,if you want to download other user files we need to enter the validate key. The key is already sent to that user through email

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

	The screen shows the email with the validity code
	The screen shows the process of accessing the other user file through their validate key

6.CONCLUSION:

In this paper, we study on how to solve the problem of user's privacy leakage in cloud storage auditing with deduplication when brute-force dictionary attacks are launched. We design a lightweight cloud storage auditing scheme with deduplication supporting strong privacy protection. In the proposed scheme, the privacy of user can be well preserved against the cloud and other parties. The user relieves the heavy computation burden for generating data authenticators and verifying data integrity. The security proof shows that the proposed scheme is secure. We also provide detailed comparisons among our proposed scheme and other existing schemes by experiments. Experimental results show the proposed scheme achieves higher storage efficiency and is more efficient in authenticator generation phase and auditing phase.

7.REFERENCE:

[1] The Gnu Multiple Precision Arithmetic Library (GMP). Accessed: Oct. 2019. [Online]. Available: http://gmplib.org/

[2] G. Ateniese, R. Burns, R. Curtmola, J. Herring, L. Kissner, Z. Peterson, and D. Song, ''Provable data possession at untrusted stores,'' in Proc. 14th ACM Conf. Comput. Commun. Secur. (CCS), 2007, pp. 598–609.

[3] M. Bellare, S. Keelveedhi, and T. Ristenpart,''Message-locked encryption and secure deduplication,'' in Proc. Annu. Int. Conf. Theory Appl. Cryptograph. Techn. Berlin Germany: Springer, 2013, pp. 296–312.

[4] H. Cui, R. H. Deng, Y. Li, and G. Wu, ''Attribute based storage supporting secure deduplication of encrypted data in cloud,'' IEEE Trans. Big Data, vol. 5, no. 3, pp. 330–342, Sep. 2019.

[5] R. Ding, H. Zhong, J. Ma, X. Liu, and J. Ning, ''Lightweight privacypreserving identity-based verifiable IoT-based health storage system,'' IEEE Internet Things J.,

vol. 6, no. 5, pp. 8393–8405, Oct. 2019.

[6] J. R. Douceur, A. Adya, W. J. Bolosky, P. Simon, and M. Theimer, ‘‘Reclaiming space from duplicate files in a serverless distributed file system,’’ in Proc. 22nd Int. Conf.

Distrib. Comput. Syst., Jul. 2002, pp. 617–624.

[7] Y. Fan, X. Lin, G. Tan, Y. Zhang, W. Dong, and J. Lei, ‘‘One secure data integrity verification scheme for cloud storage,’’ Future Gener. Comput. Syst., vol. 96, pp. 376–385, Jul. 2019.