IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, April 2024

Secure Cloud Storage with Deduplication Technique

Ms. Aishwarya S. Khutwad[1], Mr. Ajay P. Ganore[2], Mr. Shubham B. Ganore[3], Mr. Vishal P. Shinde[4]
Lecturer, Department of Information Technology[1]
Students, Department of Information Technology[2,3,4]
Mahavir Polytechnic, Nashik, Maharashtra, India

Abstract: *Data de-duplication refers to providing cloud providers with a way to manage the uncontrollable data and the challenges of cloud storage. The success of IOT and social media led us to face big data challenges. Big data is interesting but also becomes a critical challenge for cloud service providers. Data storage and management become also the topic of discussion where big data generate business opportunities as well as come with big issues for the cloud providers. In this paper, we discussed the issues of redundant data and techniques to prevent data redundancy on the cloud. A third-party auditor checks the user's data for correctness and gives the accuracy of the data that is stored in a cloud server. The communication and computation overheads were reduced. The deduplication technique is used to check whether the file that users need to store in cloud storage already exists on the cloud server.*

Keywords: Cloud service provider; deduplication; third party auditor; data dynamics

I. INTRODUCTION

Cloud storages server gives an ideal service model [1]. Cloud storage means "the storage of data online in the cloud". Different companies store their data and access data from many different and connected resources that comprise a cloud. Once the owner of the data transfers the data to the cloud storage server, he removes its copy from the local machine, thereafter owner will not be able to access the data locally. In this situation, the primary issues are the integrity and confidentiality of the data transferred. The regular issue is Data Loss Leakage, cloud providers are not trustworthy [2]. Cloud storage providers may hide data loss occurrences just to keep up with their reputation. For enhancing service-oriented accountability, the cloud server may confirm to motivate the clients that, their data is not tempered or discarded [2]. The cloud servers are not completely trustworthy. The framework for the Remote Data Possession Checking Protocol was proposed by Chen et al. [4]. The third-party auditor checks the user's data correctness and also verifies the accuracy of the data that is stored in the cloud server and this reduces the burden of cloud users.

Verification can be done any number of times without the involvement of the verifier to compare against the original data. This system will reduce communication and computation overhead. The proposed scheme also supports block-level data dynamic operations which include insert, delete, modify, search, and update. A data compression technique that eliminates duplicate copies of data blocks or files in storage is known as Data deduplication. No need to check whether the content or file is stored previously in the cloud. The system will check duplicate data available and notify the user. Hence, only one copy will be there in storage. This technique improves storage utilization. Deduplication can be done at either the block level or the file level. For file-level deduplication, replicas of the same file are eliminated and block-level replicas of the same blocks of data are eliminated that are present in non-identical files [5].

II. METHODOLOGY

- Cloud API
- Data Integration
- Encryption
- Hashing Algorithm
- Duplication Detection

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, April 2024

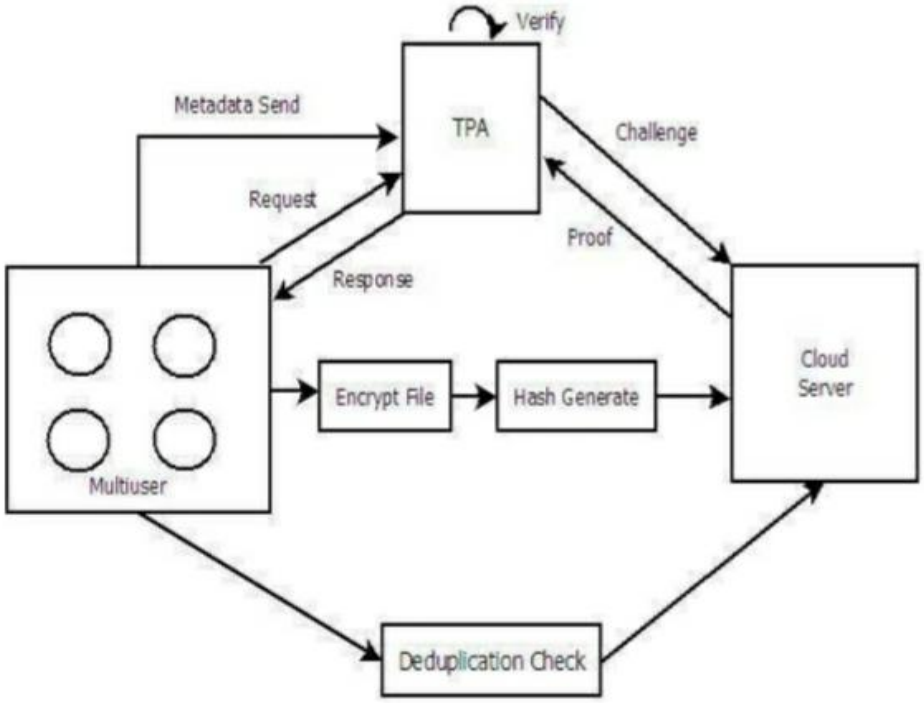


Fig.1. System Architecture

Before Dynamic RDPC protocol for cloud storage is proposed to remove security flaws, improve efficiency and reduce storage space required. In this Proposed system accuracy of the user's data is checked by Third party auditor (TPA) instead of the user. Hence, the user's computation burden will be reduced. We use a homomorphic hash function is used with the Paillier algorithm for hash key generation. Merkle Hash Tree (MHT) is used to support the data dynamics operations.

Users select the file from the local machine that he wishes to store in the cloud server. Then divide a file into blocks, encrypt the file using the Paillier algorithm, generate the hash of the file using a secure hash function, and send that file to the cloud server. If a user needs to check whether his file or data is in place or not, he sends the request to TPA. TPA in turn will generate the challenge and verify the accuracy of the file in a cloud server. TPA receives the proof from the cloud server; TPA will verify it and send the response back to the user. The response will be either true or false

III. MODELING AND ANALYSIS

The points of improved and secured dynamic RDPC protocol are given below [17]:

- Key Generation: Generates public, private, and secret keys. Tag Generation: Given a file F, the cloud client splits the file F into n different blocks and divides each block into m sectors.
- Challenge: On behalf of the user TPA generates the challenge and sends it to the server.
- Proof Generation: The server receives a challenge from the TPA. It computes the proof and sends it back to TPA.
- Proof Verify: TPA receives proof from the server and it matches to proof generated by using its metadata and returns the response to the user.
- If data in the cloud server is intact then TPA returns true to the user else returns false

Paillier algorithm

Paillier algorithm is used for key generation. The probabilistic asymmetric algorithm for public key cryptography.

Key generation

Choose any two large prime numbers arbitrarily p,q. gcd(pq, (p-1)(q-1)) =1

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, April 2024

Calculate the value of n.
n = pq and λ= lcm(p-1, q-1).
Select any integer g arbitrarily.
Make sure n divides the order of g by examining the existence of modular multiplicative inverse.

Data dynamics operations
Data dynamics is done at the block level, all data dynamics operations can be performed. When a record is modified, verification metadata is required to be changed. Hence, the required modifying operating cost should be less.

Hashing algorithm
Start
Read data owner id(avoid)
If (doid && udoid == Null)
Stop
Read file header log file data
Retrieve No. of blocks from Stored hash XML
Select the block number the user want to verify.
Get the auxiliary information for the block from the stored hash XML
Based on Auxiliary information generate a new root for NHT
If (new root ≠) data modified
Else File is not modified
Stop.

Advanced Encryption Standard (AES)
Advanced Encryption Standard (AES) is a symmetric-key block cipher algorithm and it is the standard of the United States Government's Federal Information Processing. The design and strength of all key lengths of the AES algorithm were sufficient to protect classified information up to SECRET level and TOP SECRET information (either the 192 or 256 key lengths). The key length of AES may be 128, 192, or 256 bits but it uses a fixed block size of 128 bits. The input bytes array is first mapped to the State array at the beginning of encryption or decryption. Finally, the final value of the State is mapped to the output array bytes. The number of rounds in AES depends on the length of the key. There are 10 rounds for 128-bit keys and 12 rounds for 192-bit keys in the AES algorithm. 256-bit keys will perform 14 rounds. Different 128-bit round keys are calculated from the original AES key and used in each of these rounds. Each block in AES performs four primary processes, namely SubBytes, ShiftRows, Mix Columns, and Add Round Key. The decryption of AES cipher text is the reverse operation of the encryption process and subkeys are also in reverse order. During the decryption, each round consists of the four processes conducted in the reverse order
− InvAddRoundKey, InvMixColumns, InvShiftRows, and InvSubBytes. Since the decryption is the reverse of the encryption process, the decryption algorithm is separately implemented, although they are very closely related. AES has been widely adopted in today's cryptography because of its support in both hardware and software. At present, the AES has not been broken and it can be only broken by brute force or exhaustive search. The AES algorithm was designed to make it difficult to break by linear and differential analysis. AES makes it difficult for exhaustive key searches due to having a flexible key length

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, April 2024

IV. RESULTS AND DISCUSSION

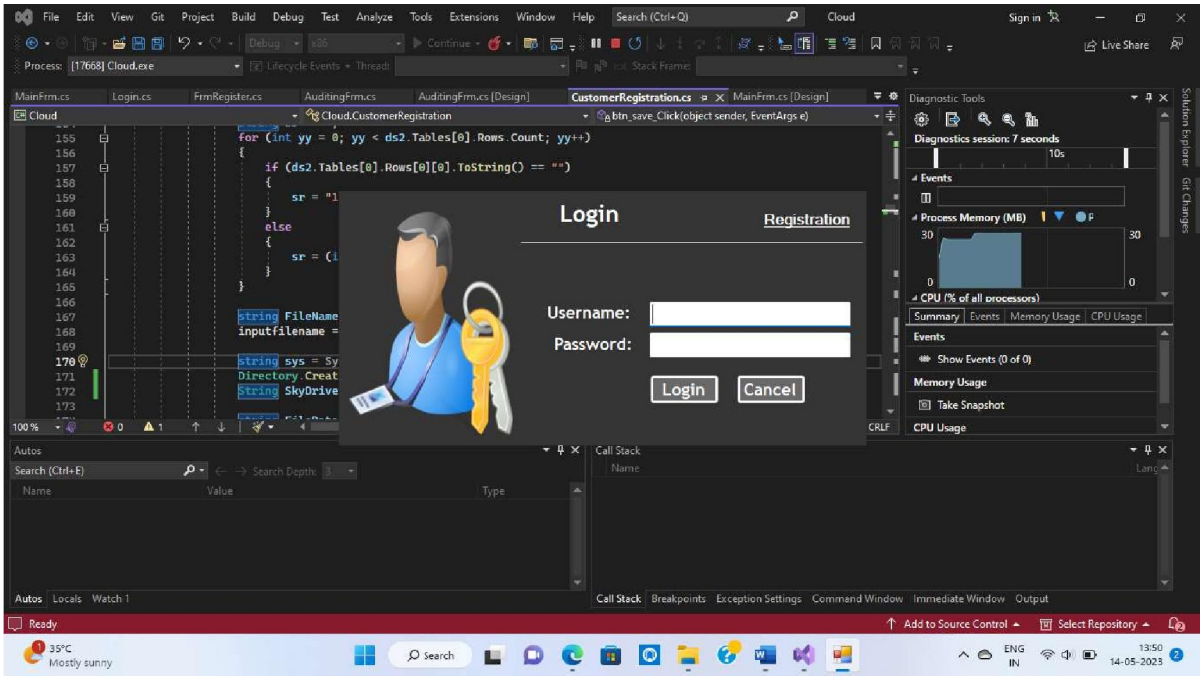


Fig.2 Login Form

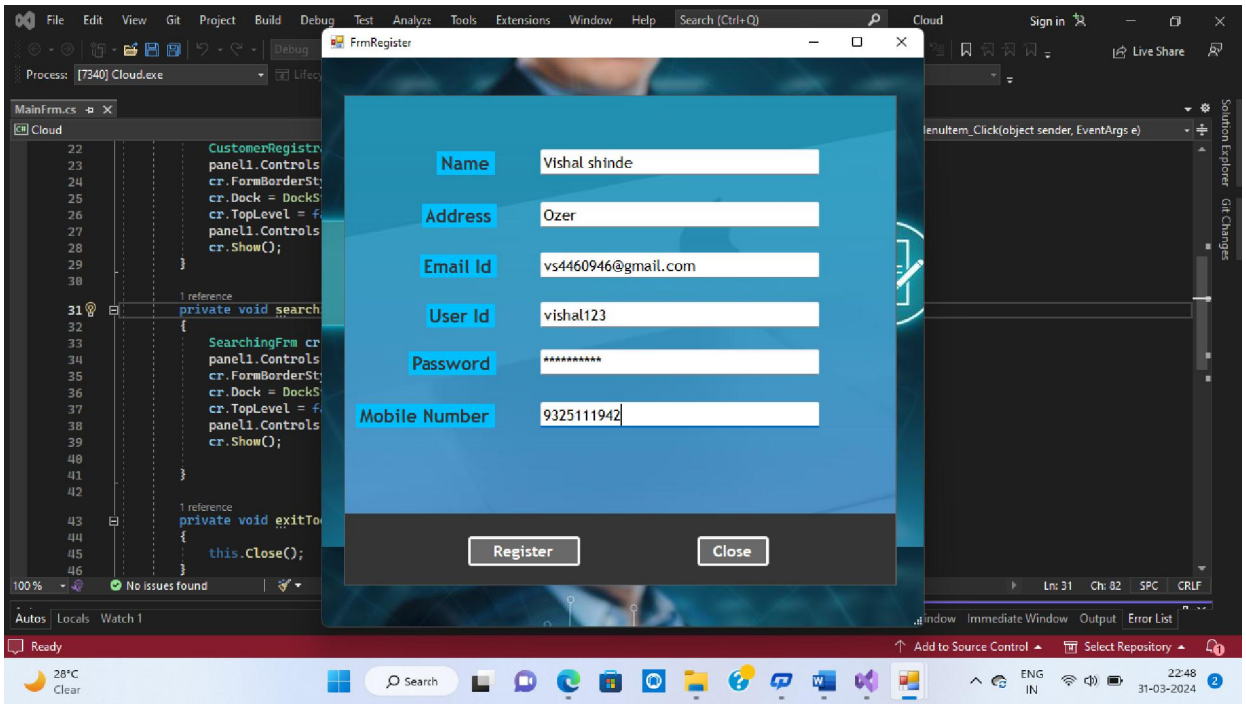


Fig.3. User Registration

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, April 2024

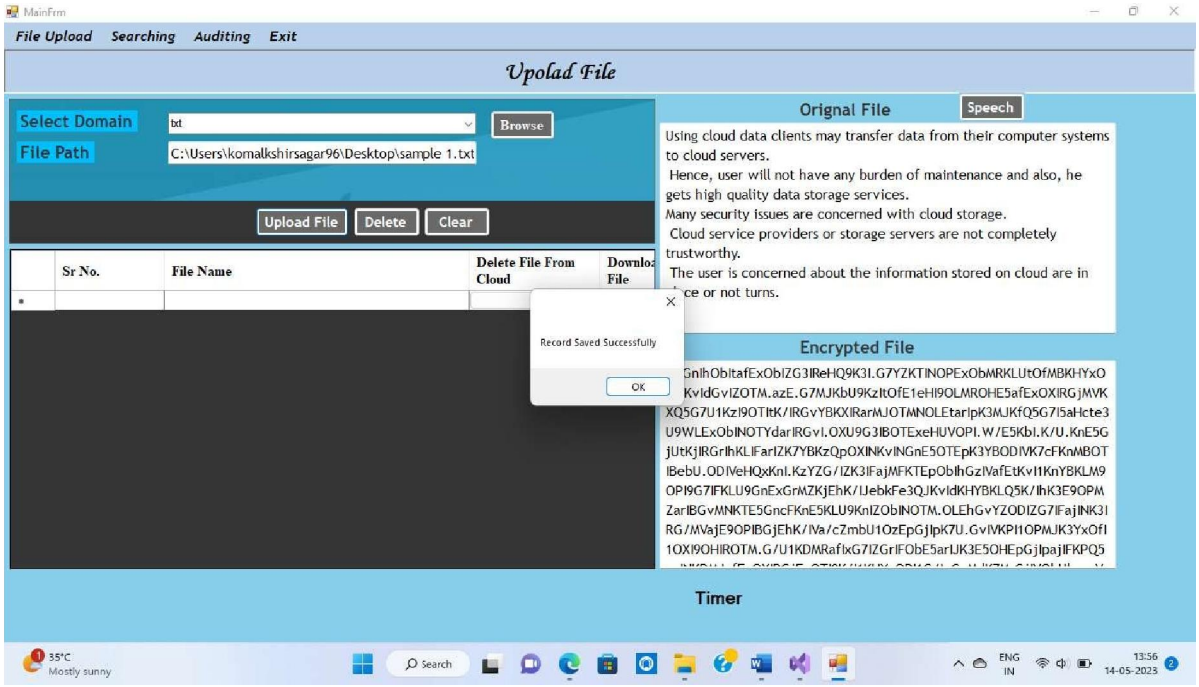
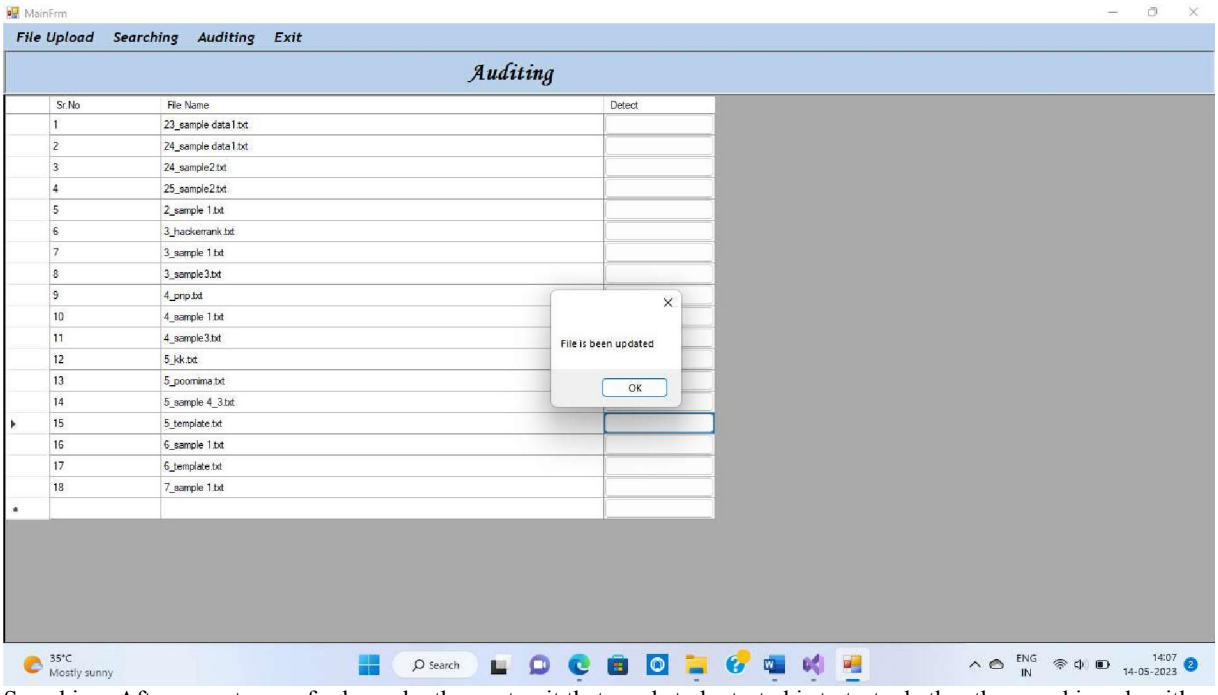


Fig.4. File Upload & Deduplication

Fig.5. Fraud Detection



Searching: After acceptance of schema le, the next unit that needs to be tested is to test whether the searching algorithm is working properly or not.

Encryption: To test whether the encryption algorithm is giving properly encrypted data as an output or not.

Fraud Detection: To check whether the hashing or signature key generation algorithm is properly working or not to validate the fraud on the cloud server.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, April 2024

Alerts generation: To validate that the proper alerts are going to the proper user. As shown in the table, different test cases are analyzed as per the requirement of the system. It gives test cases for installation.

V. CONCLUSION

The main goal of this system is to determine an improved RDPC protocol for cloud storage is proposed which checks the data integrity. This system supports block-level data dynamics including insert, update and delete. It provides integrity protection of customers' data. It also uploads a similar cloud record which will cause information losses and also expand the extra space on a cloud server, while our proposed system supports the deductibility process for checking file that exists already at the cloud server or otherwise. The existing framework does not support the deduplication process. If there's a cloud file then the same file doesn't need to be stored in the cloud again. This protocol is secure next to an untrusted server and also prevents an attack from being substituted. Third-party auditors are used to reduce the cloud burden.

VI. ACKNOWLEDGMENT

We would like to take this opportunity to express our gratitude and acknowledge the invaluable contribution of Prof. A. S. Khutwad in the development of this research paper. Their guidance, expertise, and support helped shape our research question, define the scope of the study, and develop the project methodology. Their feedback throughout the research process was crucial in helping us refine our ideas and improve the quality of our work.

We are deeply grateful for their time, effort, and dedication to our research project, and would like to express our sincerest appreciation for their mentorship and support.

REFERENCES

[1] M amdaqa, M., & Tahvildari, L. (2012). Cloud Computing Uncovered: A Research Landscape. H. Ali & M. Atif (Eds.), Advances in Computers Elsevier. 41–85.

[2] Wang W., Zeng, G., Yao, J. (2012). Cloud-DLS: Dynamic trusted scheduling for cloud computing original research article. Expert Systems with Applications, 39(3), 2321-2329.

[3] Lin Y., Chang, P. (2011). Maintenance reliability estimation for a cloud computing network with node failure. Expert Systems with Applications, 38(11), 14185-14189.

[4] Chen, L., Zhou, S., Huang, X., Xu, L. (2013). Data dynamics for remote data possession checking in cloud storage. Computers and Electrical Engineering, 39, 2413-2424.

[5] Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou. A Hybrid Cloud Approach for Secure Authorized Deduplication. IEEE Transactions On Parallel And Distributed System Vol: Pp No:99 (2014).

[6] Ateniese, G., Burns, R., Curtmola, R., Herring, J., Kissner, L., Peterson, Z., et al. (2007) Provable data possession at untrusted stores. In ACM CCS 2007, ACM, 598–609

[7] Shacham, H., Waters, B. (2008). Compact proofs of retrievability. ASIACRYPT 2008 (Vol. 5350, pp. 90–107). Berlin/ Heidelberg: Springer

[8] Ateniese, G., Pietro, R. D., Mancini, L. V., Tsudik, G. (2008). Scalable and efficient provable data possession. In SecureComm'08 (pp. 1–10)

[9] Erway, C., Kupcu, A., Papamanthou, C., Tamassia, R. (2009). Dynamic provable data possession. In ACM CCS'09 (pp. 213–222).

[10] Wang, Q., Wang, C., Ren, K., Lou, W., Li, J. (2012). Enabling public audibility and data dynamics for storage security in cloud computing. IEEE Transactions on Parallel and Distributed Systems, 22(5), 847–859

[11] R. Patil Rashmi, S. M. Sangve (2015)" Public auditing system: Improved remote data possession checking protocol for secure cloud storage", International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), 75-80.