Secure De-Duplication in Cloud Computing Environment by Managing Ownership Dynamically

Shashikala M.K
Assistant Professor
Dept. of Computer Science & Engineering
Rajeev Institute of Technology
Hassan, India

Dhruva M.S
Assistant Professor
Dept. of Computer Science & Engineering
Rajeev Institute of Technology
Hassan, India

***Abstract*—As the arena movements to superior ability for recorded purposes, the usage of Cloud Service Providers (CSPs) are increasing more ubiquity. With the steady increment of disbursed garage adopters, records de-duplication has become a want for cloud providers. In distributed storage administrations statistics de-duplication is one of crucial structures to lessen the distance requirements of administrations by way of casting off replica duplicates of rehashing facts and setting away unmarried replica of them. But it leads to security troubles when specific clients deliver indistinguishable records to the allotted storage. As of past due, a few de-duplication plans were proposed to take care of this problem. However the general public of the plans enjoy the unwell effects of protection issues, given that they don't take into account the dynamic changes inside the obligation for statistics. In this paper, a unique server-facet de-duplication plot is proposed for combined statistics that utilizations RCE and gathering key management machine. The proposed conspire ensures that one-of-a-kind authorized access to the commonplace facts is conceivable. The security of the de-duplication schemes is furnished by means of making use of suitable encryption schemes.**

Keywords-Deduplication; Cloud Providers; Encryption; Security

I. INTRODUCTION

Cloud computing lets in get entry to unlimited virtualized sources from everywhere and at any time thru the net. Now-a-days the quick development of facts volumes positioned away within the disbursed storage has brought about an elevated interest for strategies for sparing plate space and machine facts transmission. Cloud professional co-ops, for instance, Dropbox[6], Wuala[7], Mozy[8], and Google Drive[9] always search for systems meant to restrict excess data and enlarge space reserve budget. A survey found out that best 25% of the information in facts warehouses are particular. We recognition on de-duplication, that's a specialized technique to save the digital space by way of casting off multiple copies of identical records in garage. Rather than retaining various facts duplicates with a comparable substance, de-duplication distinguishes repetition in facts and later on kills excess records by means of maintaining only a single bodily replica and alluding other extra records to that duplicate.

Despite the fact that records de-duplication brings a remarkable deal of advantages, however safety and records privacy is as yet sensitive problems. Customary encryption is incongruent with statistics de-duplication. Encryption of the indistinguishable statistics duplicates of numerous clients with numerous encryption keys will set off one of a kind parent writings, which make de-duplication unattainable. The satisfactory execution of standard encryption might define be capable of as takes after: Consider clients An and B, scrambles the file M beneath their mystery keys SA and SB and stores their referring to ciphertext CA and CB. At that factor, additionally problems emerge: (1) by using what way can the cloud server experience that the vital record M is nearly identical, and (2) irrespective of the opportunity that it can distinguish this, how would possibly it empower each clients to recoup the information, in view in their exceptional mystery keys? One critical way out is to empower on every patron to encode the record with trendy key of the disbursed garage server. By then, the server can de-duplicate the distinguished development by way of unscrambling it with its personal key combine. Still, this association allows the cloud server to get the outsourced simple statistics, which may separate the security of the records if the cloud server cannot be definitely depended on. Hash table calculation defeats the drawback which examined previously. It encodes/decodes facts replica with a joined key, that's gotten through registering the cryptographic hash estimation of the substance of the information duplicate. At that factor, customers maintain the keys and give the ciphertext to the cloud server. Since the joined encryption is deterministic, indistinguishable files will create a comparable parent content.

Suppliers of cloud-primarily based storage, for instance, Google power would save be able to on restriction costs via de-duplication: should clients exchange a similar file, the management perceives this and stores only unmarried replica. Concurrent encryption has been proposed to execute information safety whilst making de-duplication manageable. It encodes/unscrambles facts replica with a united key, that's gotten via registering the cryptographic hash estimation of the substance of the facts replica. After key generation and records encryption, customers maintain the keys and ship the determine content to the cloud.

DOI:10.24032/ijeacs/0206/05

II. Literature Survey

Literature review is the process of presenting the summary of the conference papers and journal articles study resources. The information de-duplication conspires over encoded information has been produced and enhanced further into Convergent Encryption (CE), Leakage-Resilient (LR) De-duplication plot, Randomized Convergent Encryption (RCE) and Dynamic Ownership Management Scheme.

A. Merged Encryption (CE)

LI[1] With a particular true objective to keep information security against inside cloud server and furthermore outside challengers, clients may require their data encoded. Be that as it may, customary encryption under various clients' keys makes cross-customer de-duplication incomprehensible, since the cloud server would constantly watch unmistakable ciphertexts, regardless of the possibility that the information are the same, paying little respect to whether the encryption calculation is deterministic. Douceur [2] presents Convergent Encryption, which is the promising answer for this issue. In CE, an information proprietor gathers an encryption key over information by using cryptographic hash work. At that point registers the ciphertext utilizing square figure over information alongside their encryption key. CE deletes information and keeps just encryption key subsequent to transferring ciphertext to the distributed storage. Since encryption is deterministic, on receipt of same record CE delivers same ciphertext for it and the server does not store the record yet rather refreshes meta-information to show it has an extra proprietor.

Benefits: Provides promising arrangement over ordinary encryption and jelly information security.

Restrictions: Convergent Encryption experiences some security issues i.e. label consistency issue. It implies that honesty and security of information has been traded off because of the absence of PoW process and dynamic possession administration.

B. Ramp Secret Sharing Scheme (RSSS)

Li [3] formalizes a joined key administration plot i.e. Dekey which is proficient and solid for secure de-duplication. Dekey set de-duplication between united keys and circulates those keys over various key servers while saving the semantic security of joined keys and protection of outsourced data. Dekey is actualized utilizing the Ramp mystery sharing plan. Dekey utilizes RSSS to gather concurrent keys. It's thought is to allow de-duplication in joined keys and circulate the merged keys over different KM-CSPs. Rather than scrambling the joined keys on a for every client premise, Dekey manufactures mystery shares on the first merged keys (that are in plain) and appoints the offers over different KM-CSPs.

Benefits: Provides dependable, proficient and adaptation to internal failure united key component for secure de-duplication.

Constraint: This plan does not bolster dynamic proprietorship administration issue in secure de-duplication.

C. Spillage Resilient (LR) De-duplication Scheme

Xu[4] proposes a spillage versatile de-duplication plan to unravel the information respectability issue. It tended to a key security worry in cross-client customer side de-duplication of encoded documents in the distributed storage: protection of clients' delicate records against both outside challengers and the honestbut-inquisitive distributed storage server in the limited spillage display

Benefits: Resolves information trustworthiness issue i.e. anticipates label consistency assault.

Restrictions: Secure verification of possession (PoW) plot in the standard model remains an open issue. Another downside is the absence of dynamic possession administration among the information holders.

D. Approved De-duplication Hybrid Cloud

LI [5] proposes an approved de-duplication conspire where differential benefits of clients, and in addition the information, are considered in the de-duplication strategy in a cross breed cloud condition. He exhibited a few new de-duplication developments supporting approved copy check in cross breed cloud design, in which the copy check tokens of documents are created by the private cloud server with private keys. The figure demonstrates the engineering of approved de-duplication.

Benefits: This plan gives approved de-duplication over half and half cloud for clients who have distinctive benefits.

Restrictions: Data spillage.

Far site – a de-duplication framework that spotlights less on approval of clients. Information on the cloud is put away in typical frame. De-duplication is performed just on content and pictures and it does not bolster all document formats [10]. In 2013, Neal Leavitte[4] examined the diverse issues landing in the de-duplication in multi-inhabitant condition. Distinctive creators proposed the utilization of the single key encryption.

III. Existing System

In existing framework, Cryptographic methods were connected to get to control for remote stockpiling frameworks. The information proprietors scramble records by utilizing the symmetric encryption approach with content keys and after that utilization each client's open key to encode the substance keys. It requires every information proprietor to be online constantly. A few strategies convey the key administration and circulation from the information proprietors to the remote server under the suspicion that the server is trusted or semi-trusted.

Disadvantages

- The key management is very complicated when there are a large number of data owners and users in the system.
- The key distribution is not convenient in the situation of user uses the system dynamically.

- The user needs to know private key.
- Less protect security.

These de-duplication systems cannot support differential authorization duplication check.

IV. PROPOSED SYSTEM

We propose a de-duplication plot over scrambled information. The proposed conspire depends on Elliptic Curve Cryptography and declines the key length while giving securities at an indistinguishable level from that of different cryptosystems gives. The proposed plan ensures that solitary affirmed access to the regular data is possible. We give the unusual state security and avoid the replication of archive in the cloud expert association. To secure the assurance of tricky data while supporting de-duplication, the centered encryption methodology has been proposed to encode the data before outsourcing .In proposed framework, we are utilizing hash capacity to produce key for the record. By utilizing hash capacity to keep away from the duplication in cloud. After that we are applying cryptographic system for security reason. We are utilizing ECC calculation for encryption and decoding process. The proposed scheme has the going with purposes of enthusiasm in regards to security and profitability:

To begin with, dynamic proprietorship organization guarantees the retrogressive and forward secret of de-duplicated data upon any ownership change. Second, the proposed plot ensures security in the setting of PoW by showing a re-encryption framework that uses an additional social event key for dynamic ownership gathering.

A. *Secure Data De-duplication Architecture*

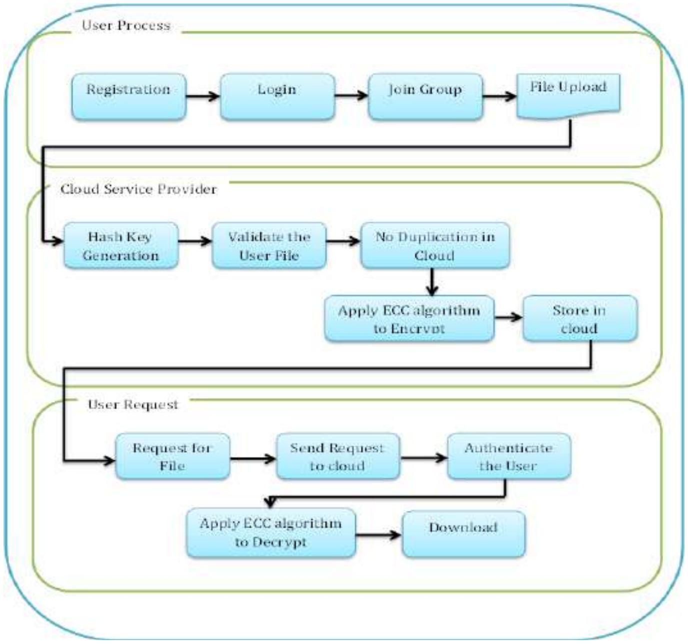


Figure 1

B. *Sequence Diagram*

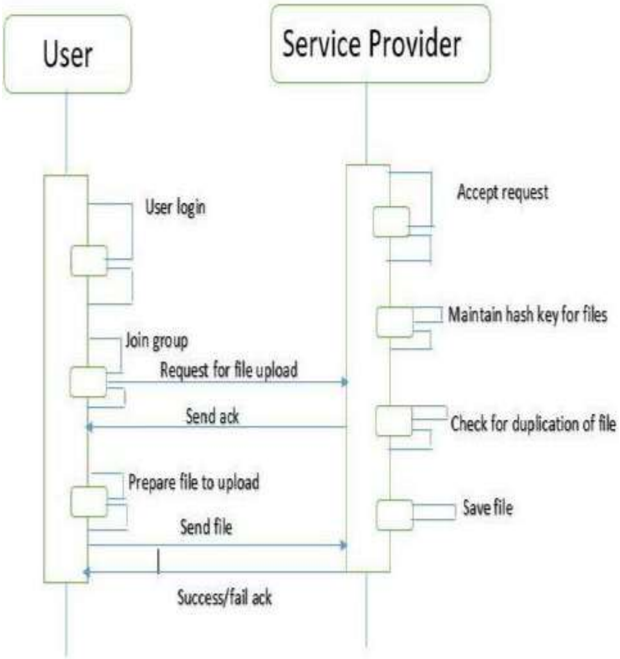


Figure 2.

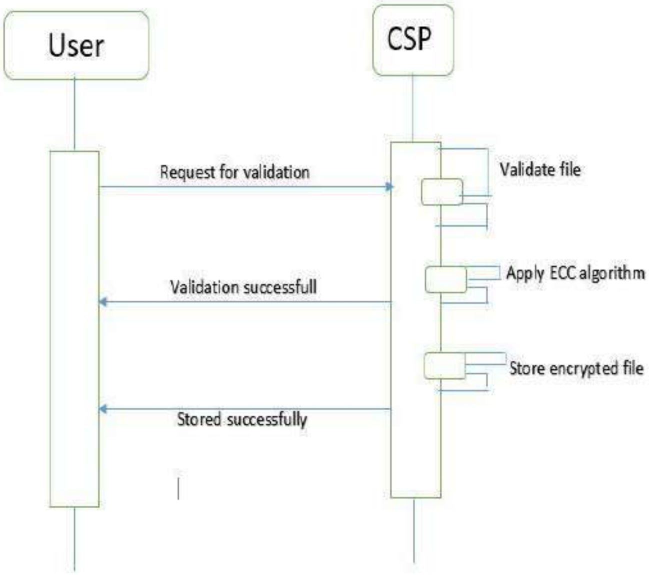


Figure 3. User interaction with Service Provider

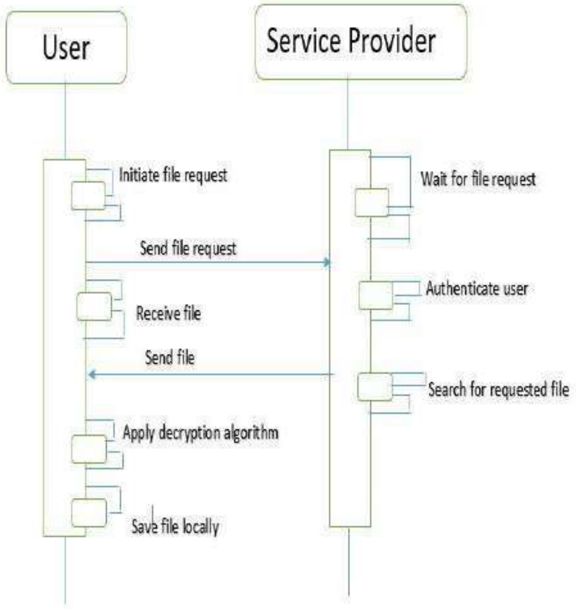


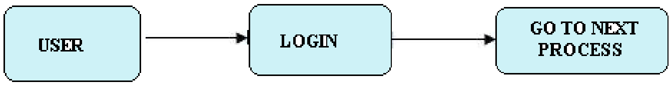
Figure 4. User interaction with Service Provider

V. METHODOLOGY

MODULES:

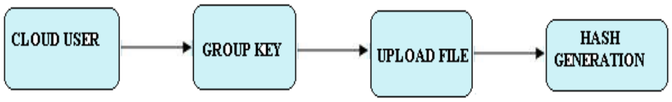
1) User Registration and Login

In this Module If he is a new user he needs to enter the required data to register the form by providing the user details like name; dob etc. and the data will be stored in server for future authentication purpose. After registration user will get the username and password for further process. Using Username and Password, user login into Group. Group generate key for the valid user and process inside the group under the valid key.



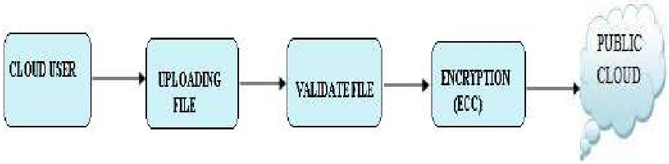
2) User Joining The Group and File Upload

For every user a key would be generated using which the user gets the authentication to join the group with a key. In file upload process, user choose the file from the system and generate hash key for each file. Hash key generation is provided to avoid duplication of the file to the cloud. If the file is already in the cloud the user cannot upload the file.



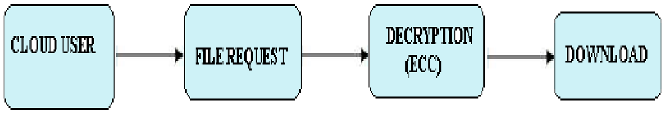
3) File Encryption and Storage in Cloud

If data duplication check is negative, the data user encrypts its data using ECC algorithm in order to ensure the security and privacy of data, and stores the encrypted data at CSP. We implement ECC algorithm which converts a file in to a binary format and it gets encrypted and is stored on to the cloud. The data that is stored on to the cloud will be in encrypted format.

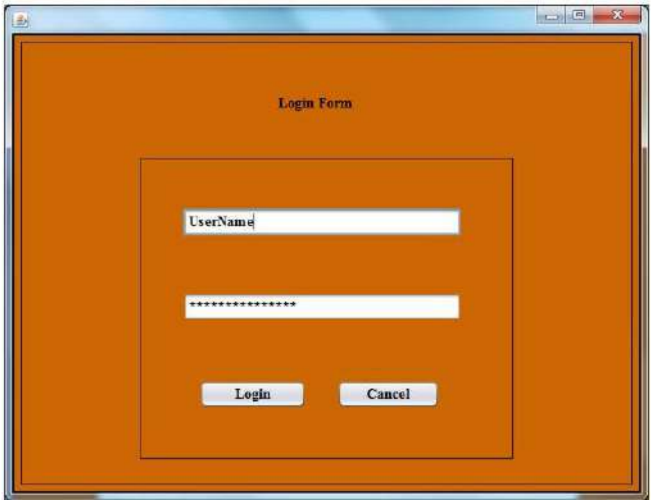


4) User File Request and Download

Any user who has registered earlier and joined the group with a valid key can request the file to the cloud. The cloud service provider after authenticating the user can receive the file request, decrypt the file using ECC algorithm and send the requested file to the user. Then the file will be downloaded in the user's location.



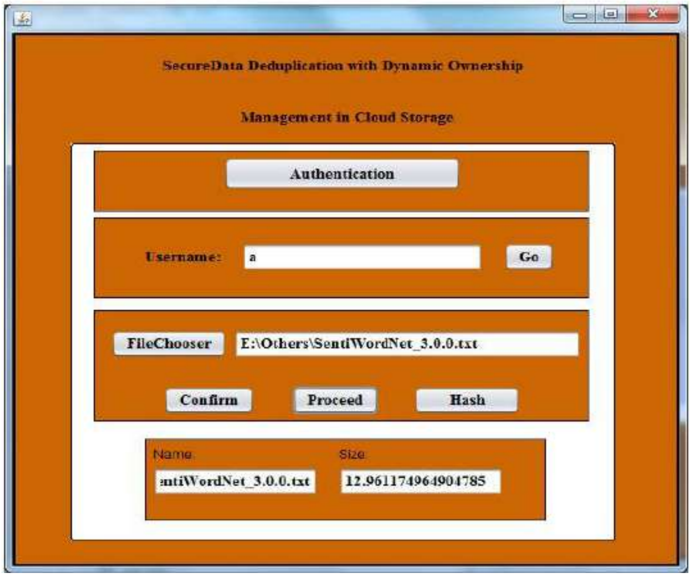
VI. RESULTS


Login Form

UserName

****

Figure 5. Login Page



SecureData Deduplication with Dynamic Ownership

Management in Cloud Storage

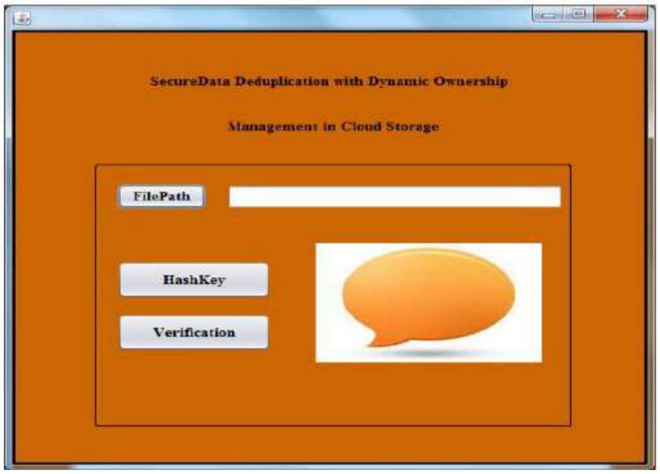
Username: a

FileChooser: E:\Others\SentiWordNet_3.0.0.txt

Name: ntiWordNet_3.0.0.txt

Size: 12.961174964904785

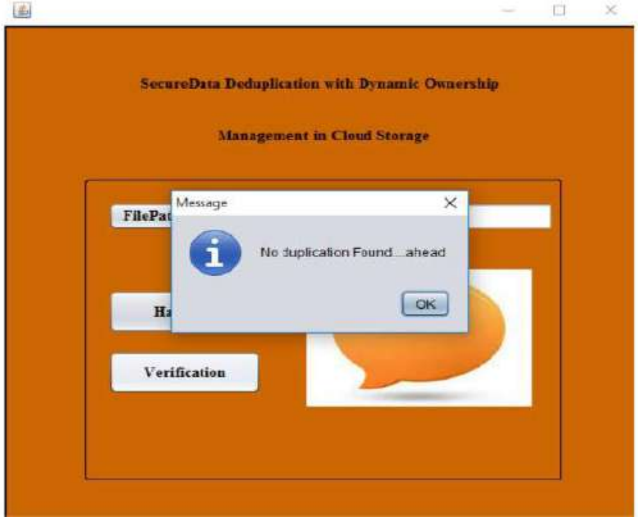
Figure 6. Uploading the file



SecureData Deduplication with Dynamic Ownership

Management in Cloud Storage

Figure 7. Generating hash key and verifying file



SecureData Deduplication with Dynamic Ownership

Management in Cloud Storage

Message

No duplication Found....ahead

Figure 8. Checking Duplication of File

VII. CONCLUSIONS

Managing encrypted facts with de-duplication may be very critical and important in examine for achieving a prospering cloud storage provider, mainly for massive data storage. . In this paper, we proposed a realistic scheme to control the encrypted big facts in cloud with de-duplication based on possession venture. This plan enhancements facts safety and privacy in disbursed garage towards any clients who don't have good sized responsibility for data, and also in opposition to an actual however inquisitive cloud server. By utilizing hash key successfully can accomplish de-duplication in dispensed storage. Elliptic curve cryptography is applied for encryption and decryption method to decrease probability of attacking the document.

REFERENCES

[1] J. Li, X. Chen, X. Huang, S. Tang, Y. Xiang, M. Hassan, and A. Alelaiwi, “Secure Distributed Deduplication Systems withImproved Reliability,” IEEE Transactions on Computer, Vol. 64, No. 2, pp. 3569–3579, 2015

[2] R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M.Theimer, “Reclaiming space from duplicate files in a server less distributed file system,” Proc. International Conference on Distributed Computing Systems (ICDCS), pp. 617–624, 2002.

[3] Li, X. Chen, M. Li, J. Li, P. Lee, and W. Lou, “Secure deduplication with efficient and reliable convergent key management,” IEEE Transactions on Parallel and Distributed Systems, Vol. 25, No. 6, 2014.

[4] Xu, E. Chang, and J. Zhou, “Leakage-resilient client-side deduplication of encrypted data in cloud storage.”

[5] J. Li, Y. K. Li, X. Chen, P. Lee, and W. Lou, “A hybrid cloud approach for secure authorized deduplication,” IEEE Transactions on Parallel and Distributed Systems, Vol. 26, No. 5, pp. 1206–1216, 2015.

[6] Dropbox, http://www.dropbox.com/.

[7] Wuala, http://www.wuala.com/.

[8] Mozy, http://www.mozy.com/.

[9] Google Drive, http://drive.google.com

[10] Rev De dup: A reverse deduplication storage system optimized for reads to latest backups 2013.

AUTHORS' PROFILE

Shashikala M. K completed her B.E. degree and M. Tech. degree in Computer Science and Engineering from Visvesvaraya Technology University, Belgaum, India Currently she is working as Asst. Professor in the Department of Computer Science and Engineering at Rajeev Institute of Technology, Hassan, India. Her areas of interest include networks, algorithms.

Dhruva M.S. completed his B.E. degree and M.Tech. degree in Computer Science and Engineering from Visvesvaraya Technology University, Belgaum, India. Currently he is working as Asst. Professor in the Department of Computer Science and Engineering at Rajeev Institute of Technology, Hassan, India. His areas of interest include multimedia networks, Compiler Design and Algorithms.