ATLANTIS PRESS

Advances in Engineering Research (AER), volume 142

International Conference for Phoenixes on Emerging Current Trends in Engineering and Management (PECTEAM 2018)

Cloud Secure Distributed Storage Deduplication Scheme for Encrypted Data

[1]N. Indira, [2]Dr. S. Rukmani Devi
[1]*Research Scholar,* Department of Computer Science and Engineering,
College of Engineering, Anna University, Chennai, India
[2]Professor, Department of Computer Science and Engineering, R.M.D. Engineering College, Kavaraipettai, India,
indiranatarajan13@gmail.com, rdrukmani319@gmail.com

Abstract— **Deduplication portrays the end of copy or excess data and it expels the monotonous data previously putting away it. These strategies are generally utilized for information reinforcement, arrange minimization, and capacity overhead. Since a long time ago settled deduplication plans have limitations on encoded information and security. The principle point of this task is to accomplish new dispersed de-duplication frameworks with higher unwavering quality in which the information lumps are appropriated crosswise over HDFS and solid key administration system is utilized for secure de-duplication utilizing slave hubs. Rather than having various duplicates of a sole substance, deduplication expels repetitive information by keeping just a single physical duplicate and proposing other excess information to that duplicate. Everyone can be elucidated in light of non-identical granularities which might be whichever way an entire record or an information square. MD5 and 3DES calculation fortify the procedure. The approach proposed here is Proof of possession (POF) of the document. Deduplication can now, legitimately address the unwavering quality and label consistency issue in HDFS stockpiling frameworks. The proposed framework prevailing with regards to diminishing the cost and time of transferring and downloading with storage room.**

Keywords: **Deduplication, monotonous, HDFS.**

I. INTRODUCTION

Cloud Computing offers services to users by revising different resources such as computational and storage services and giving them to clients in light of their resources. Cloud computing gives a very big resource pool by connecting network assets together. Data storage is the most vital and prevalent service provided by cloud. Users transfer individual or confidential information to the server formof a Cloud Service Provider and permit it to keep up these information. The same or diverse clients may transfer same content encrypted form to Cloud Service Provider, particularly for situations where information is shared among numerous clients.

Deduplication techniques are widely employed to backup data and minimize network and storage overhead by detecting and eliminating redundancy among data. Rather than of keeping several same data copies, deduplication dispenses them by keeping just a single physical copy and referring other redundant data to that copy. Each such duplicate copy can be characterized based on different granularities: it may be the entire content of the file (i.e., file level deduplication), or data block (i.e., block-level deduplication).Apart from normal encryption and decryption process.However, the method of de-duplication is executed only on single server setting.

Traditional de-duplication scheme is not scalable with the enormous increasing cloud users. Efficient key management scheme was not maintained so as to manage the convergent keys generated if there are several cloud users. We have to spend more cost for more content storage and also it is very difficult to manage more keys. Security breaches as the technique is approached over a single server setting, once hacked the information can be collected at a common server. Already existing approach is unreliable because it requires each user to securetheir master key. The existing system introduces a scheme to de-duplicate the data in cloud based on ownership challenge and proxy re-encryption.

Various customers may have similar data copies; they ought to have their own arrangement plan of restricted keys so that the same customers can get to their information or records. Specifically, every customer must member an encrypted key with each block of its outsourced mixed data copies, so as to later re-build up the data duplicates. As a result the amount of combined keys being given straightforwardly scales the quantity of blocks being secured. The gauge approach is deceitful, as it requires each customer to guarantee his own specific master key.

If the master key is accidently lost or compromised by attackers, then the user data cannot be recovered. In this paper, we propose deduplication and distributed storage of the data across the cloud using Proof Of Ownership of the file. We use Triple DES technique as the plain text is encrypted triple times with the convergent key so that data will be secured. Scalability increases as the achieved efficiently.

II. LITERATURE SURVEY

An efficient and reliable convergent key management scheme applies deduplication among convergent keys and dekey was introduced. Convergent key shares are distributed across multiple key servers, while preserving semantic security of convergent keys and confidentiality of outsourced data [5].

[3] Proposes a novel scheme based on techniques including polynomial-based authentication tags and homomorphic linear authenticators. It validates the efficiency and scalability of the scheme but still is inherently slow and impractical for bulk encryption. [1] Suggests a new component to implement the key management for each block, and makes sure that the overhead is minimal. There is no proper solution for optimization in terms of storage, bandwidth. The typical

operations like delete and edit is not mentioned.In this scheme client needs to access little amount of original file. Randomly chosen indices of the file was introduced. It reduced the burden of the client, insures remote data integrity and also fails at server side where more loads are taken [4].

A hybrid cloud approach is a combined form of private and public clouds therefore the advantage of both hit to the center of the hybrid cloud. But the flaw detected is that maintenance of Tokens for each user incurs maximal overhead [2].[8] Counteracts data leakage exclusively to disavowed clients despite the fact that they beforehand possessed that information, additionally to a legitimate yet inquisitive distributed storage server.

[9] Provides dynamic space optimization in private cloud storage backup as well as increase the throughput and de-duplication efficiency. There are no details about the various operations such as edit and delete operations.

III. PROBLEM STATEMENT

The system contains three entities: 1. Data owner (Ui) who uploads, downloads the data and is willing to delete the data that is stored into the cloud server 2. Cloud Server which provides the data storage services3. Key management server which stores and manages the keys associated with the storage.

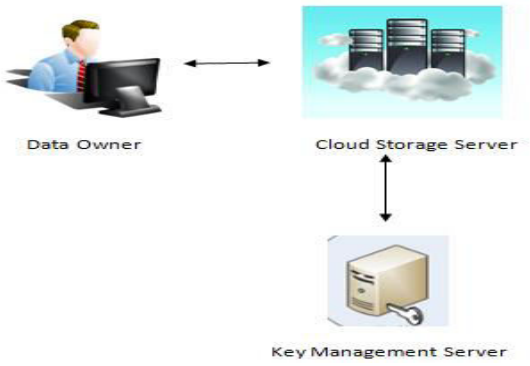

Fig.1. System Model

IV. PROPOSED SYSTEM

The proposed system depicts as the cloud user uploads the original file, at the initial phase a tag is generated to the entire file using MD5 algorithm. At the further phases the chunked blocks of each file are hashed. The convergent keys are generated respectively and the file gets uploaded to the Cloud server as cipher text. As the tag is generated for each file, the de-duplication check (both block level and file level) is done using Comma Separated Values (CSV).

Key administration slave checks copy duplicates of focalized (convergent) keys in KMCSP. Key Management slave keeps up CSV document to check evidence of verification and store keys safely. If a copy is found for the file uploaded, only the reference is generated with the soul content.

The references are stored into the slave machines, and the cloud sever will have the soul content as cipher text. As on with the download request, proof of verification (POF) for each authenticated user is verified in the cloud storage. The cloud storage supports back with the cipher text to the authorized user. With the keys generated the cipher text is decrypted and original file is given to the cloud user. Regarding the delete and edit option, only the references are deleted and edited and not the soul content.

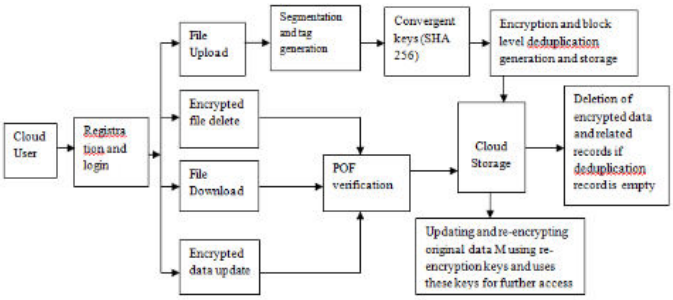

Fig. 2. System Architecture

The proposed system mainly includes four phases namely,

- Preparing Tag for upload
- Segmenting the file and storing into cloud
- Key sharing
- Decryption

Preparing Tag for upload

Clients register to the Cloud server with their data and login the page for transferring the document. User selects the file for upload to cloud server,the Cloud server stores the file and file level deduplication is checked. Tag is generated using message-digest (MD5) algorithm for producing the required hash value.

Segmenting the file and storing into cloud

We produce united keys for each block separately for block level de-duplication. Here we give a filename and secret key for file approval in future. The individual blocks are encrypted by 3DES algorithm. The original text is encoded three times with convergent keys. So when decoding the original content it requires thesame key.

Key Sharing

After encryption the convergent keys are securely presented to the key management systems. Key organization slave checks duplicate copies of centered entries in KMCSP. Key Management systems keeps up CSV archive to check confirmation of verified and stored keys securely. The customersshare the basic keys are suggested by their own specific proprietorship (proof of ownership). In the event that User required to do erasure, absolutely he should demonstrate evidence of possession to delete his own contents.

ALGORITHM

Step 1: User U_i generates data token using MD5 of original data M.

Step 2: Data token sent to CSP.

Step 3: CSP verifies token and checks the existence of duplicated data in Cloud Server.

Step 4: If file already not exists, it requests data to upload. Encrypt data using 3DES.Store M to Cloud Server.

Step 5: If already exists then generate reference with soul content. CSP re-encrypts E and provides re-encrypted key to U_j

Step 6: $U_j$can get DEK with the secret key and affirms the accomplishment of data deduplication information in the system after getting this notification.

Decryption

This download ask for necessities legitimate ownership check of the record and asserts existing tag of the client which is starting at now produced by md5 algorithm. The ownership is confirmed with the unique tag. After the authentication, the original data is decrypted by producing the request to Cloud server. Finally the original content is downloaded from each blocksin the distributed storage. The delete request made by the client will erase just the reference of the substance shared by the regular clients and not the entire substance.

V. EXPERIMENTAL RESULTS

Hybrid encryption is provided by MD5 and Triple Data Encryption Standard (3DES). MD5 produces a 128 bit hash value and specified as Internet Standard RFC 1321

MD5 Algorithm:

Step 1: Pad message so its length is 448 mod 512.

Step 2: Append a 64-bit length value to message and generate a message with 512 bits in length.

Step 3: Initialize 4-word (128-bit) MD buffer (A,B,C,D)

Step 4: Process message in 16 word (512-bit) blocks

Step5: Output is the final hash code.

Triple Data Encryption Standard (3DES) In cryptography, Triple DES (3DES), officially the Triple Data Encryption Algorithm (TDEA or Triple DEA), is a symmetric-key block cipher, which applies the Data Encryption Standard (DES) cipher algorithm three times to each data block. Three key 3DES has an thekey length of 168 bits. The screen shots given below shows the operations of Uploading of files to the Cloud, Cloud setup for file distribution and storage and encrypted data storage in the Cloud.

Fig. 3. Uploading files to cloud

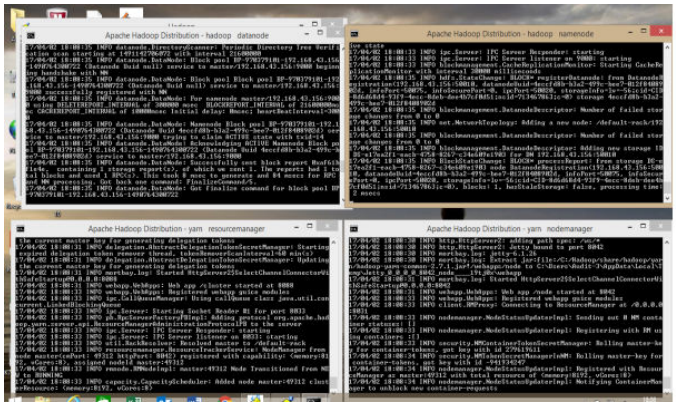

Fig. 4. Cloud Setup for file distribution and storage

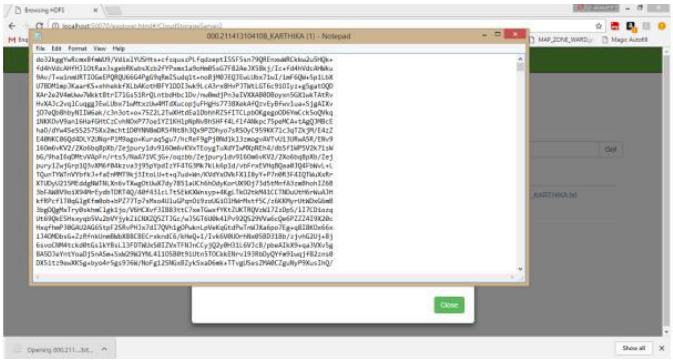

Fig.5. Encrypted storage of user content

VI.CONCLUSION

The proposed framework was tried for efficiency as far as calculation and correspondence cost prerequisites and it was discovered that our proposed strategy brought about less calculation and correspondence costs. Further, the transferring of cloud client's document to the Cloud server stockpiling was scrambled utilizing the 3DES technique which required less cost for setting up and is viewed as the most productive for our application since any endeavor to hack the framework is extremely costly and is subsequently kept away from. The Proof of proprietorship conspire was likewise solid to validate each cloud client, and the label age of each document was created utilizing MD5 calculation. Hence our framework is found to emanate elite and productivity.

REFERENCES

[1] Pasquale puzio, RefikMolva, Melekonen, “Cloud Dedup - secure deduplication with encrypted data for cloud storage”, in IEEE 5th International Conference, Dec 2013.

[2] Jagadish, Dr.SuvurnaNandyal, “A Hybrid cloud approach for secure authorized de-duplication”, published in International Journal of Science and Research (IJSR), 2013.

[3] Jiawel Yuan, Shuchengyu, “Secure and constant cost public cloud storage auditing with Deduplication”, in IEEE Conference, published in communication and network security, 2013.

[4] Chao Yang, Jianren, Jianfengma, “Provable ownership of file in deduplication cloud storage”, published in Global Communication Conference, Dec 2013.

[5] J.Li, X.Chen, M.Li, J.Li, P.Lee, and W.Lou, “Secure Deduplication with Efficient and Reliable Convergent Key Management, in IEEE Transactions on Parallel and Distributed systems, 2013.

[6] J.Stanek, A.Sorniotti, E.Androulaki, and L.Kenel, "A Secure data deduplication scheme for cloud storage", in Technical Report, 2013.
[7] C.Ng and P.Lee. Revdedup, "A reverse deduplication storage system optimized for reads to latest backups", in Proc of APSYS, Apr 2013.
[8] JunbeomHur, Dongyoung Koo, Youngjoo Shin, and Kyungtae Kang, "Secure Data Deduplication with Dynamic Ownership Management in Cloud Storage", in IEEE Transactions on Knowledge and Data Engineering, June 2016.
[9] M. Shyamala Devi, V. VimalKhanna, A. Naveen Bhalaji, "Enhanced Dynamic Whole File De-Duplication (DWFD) for Space Optimization in Private Cloud Storage Backup", in International Journal of Machine Learning and Computing, April 2014.
[10] Zhen Yan, Wenxiu Ding, Robert.H.Deng, "De-duplication on encrypted big data in cloud", in IEEE transaction on big data Vol.2, No.2, April – June 2016.