ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

A Secure Authorized De-Duplicate Checking in a Hybrid Cloud Architecture

Maddirala Venkateswarlu
M.Tech,
Department of CSE,
Newton's Institute of Engineering,
JNTUK Macherala, Guntur, AP, India.

Mr.K.Venkata Ratnam, M.Tech
Associate professor,
Department of CSE,
Newton's Institute of Engineering,
JNTUK Macherala, Guntur, AP, India.

Abstract:

A survey on industry trends has been noted where the usage of hybrid cloud architecture can be used which supports, the upcoming industry challenges by providing the efficient way of storing their data in the cloud environment by using the combination of both public and private clouds, So that it provides the facility to store sensitive data in private cloud and less critical data on to the public cloud where huge savings can be made. Since the demand for data storage is increasing day by day and by the industry analysis we can say that digital data is increasing day by day, but the storage of redundant data is excess which results in most of the storage used unnecessary to keep identical copies. So the technology de-duplication is introduced to efficiently utilize the cloud storage system. Information de-duplication is one of vital information pressure strategies for dispensing with copy duplicates of rehashing information, and has been broadly used as a part of distributed storage to decrease the measure of storage room and spare data transmission.

To secure the classifieds of touchy information while supporting de-duplication, the concurrent encryption procedure has been proposed to encode the information before outsourcing. Not quite the same as customary de-duplication frameworks, the differential benefits of clients are further considered in copy check other than the information itself. We additionally show a few new de-duplication developments supporting approved copy weighs in a half and half cloud structural engineering. To better make sure information security, this paper makes the first attempt to formally address the issue of approved information De-duplication. Security examination shows that our plan is secure as far as the definitions determined in the proposed security model. As a proof of idea, we actuality a model of our proposed approved copy check plan and behavior test bed analyses utilizing our model. We prove that our proposed approved copy check plan acquires insignificant overhead contrasted with ordinary operations.

Keywords:

De-duplication;authorizedduplicatecheck;confidentiality; hybrid cloud;convergent encryption.

I.INTRODUCTION:

In distributed computing information de-duplication is an essential information pressure part for decreasing indistinguishable duplicates of same information. this part is utilized to enhance powerful use of storage room furthermore connected to minimize information transmission over system in de-duplication strategy indistinguishable information are discover and put away amid procedure of examination. As procedure precedes other information are coordinated to the put away duplicate and at whatever point coordinated discovered the indistinguishable information is supplanted with a little reference that tended to put away information. A mixture cloud is a mix of private cloud and open cloud in which the information which is most discriminating that lives on a private cloud and the information which is effortlessly available is dwells on an open cloud half and half cloud is useful for dependability, extensibility and quick organization and price sparing of open cloud with more security with private cloud .

The unpredictable test of distributed storage or distributed computing is the game plan of vast volume of information duplication is a procedure of disposing of copy information in de-duplication strategies excess information evacuated leaving single occurrence of the information to be put away. In the past old framework the information is encoded back to outsourcing it on the cloud or system. This encryption obliges most extreme time and in addition storage room prerequisite to encode the information if there is huge measure of information around then encryption procedure gets to be mind-boggling and discriminating. By utilizing de-duplication method as a part of half and half cloud the encryption procedure get to be less complex.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

As we all realizes that the system has real measure of information which being shared by many clients. Numerous extensive systems use information cloud to store the information and offer that information on the system.As distributed computing gets to be main, an expanding measure of information is being put away in the cloud and imparted by clients to indicated benefits, which characterize the entrance privileges of the put away information. One basic test of distributed storage administrations is the administration of the constantly expanding volume of information. To make information administration adaptable in distributed computing, de-duplication has been a surely understood method and has pulled in more consideration as of late. Information de-duplication is a particular information pressure system for wiping out copy duplicates of rehashing information away. The system is utilized to enhance stockpiling usage and can likewise be connected to network information exchanges to diminish the measure of bytes that must be sent.

Rather than keeping various information duplicates with the same substance, de-duplication dispenses with excess information by keeping stand out physical duplicate and alluding other repetitive information to that duplicate. De-duplication can happen at either the record level or the peace level. For document level de-duplication, it disposes of copy duplicates of the same record. De-duplication can likewise happen at the piece level, which wipes out copy squares of information that happen in non-indistinguishable records. Distributed computing is a developing administration display that gives processing and stockpiling assets on the Internet. One appealing usefulness that distributed computing can offer is distributed storage. People and undertakings are regularly needed to remotely document their information to stay away from any data misfortune on the off chance that there are any equipment/programming disappointments or unforeseen catastrophes.

Rather than acquiring the required stockpiling media to keep information reinforcements, people and endeavors can basically outsource their information reinforcement administrations to the cloud administration suppliers, which give the fundamental stockpiling assets to have the information reinforcements. While distributed storage is appealing, how to give security assurances to outsourced information turns into a rising concern.One noteworthy security test is to give the property of guaranteed erasure, i.e., information documents are for all time difficult to heaps of cancellation.

Keeping information reinforcements for all time is undesirable, as delicate data may be uncovered later on as a result of information break or incorrect administration of cloud administrators. In this way, to keep up a strategic distance from liabilities, endeavors and government offices generally keep their reinforcements for a limited number of years and solicitation to erase (or decimate) the reinforcements a short time later. Case in point, the US Congress is detailing the Internet Data Retention enactment in approaching ISPs to hold information for a long time, while in United Kingdom, organizations are obliged to hold wages and pay records for a long time.

II.LITERATURE SURVEY:

A.Fast and secure laptop backups with encrypted de-duplication:

Many people now store large quantities of personal and corporate data on laptops or home computers. These often have poor or intermittent connectivity, and are vulnerable to theft or hardware failure. Conventional backup solutions are not well suited to this environment, and backup regimes are frequently inadequate. This paper describes an algorithm which takes advantage of the data which is common between users to increase the speed of backups, and reduce the storage requirements. This algorithm supports client-end per-user encryption which is necessary for confidential personal data.

B.Message-locked encryption and secure de-duplication:

We formalize a new cryptographic primitive, Message-Locked Encryption (MLE), where the key under which encryption and decryption are performed is itself derived from the message. MLE provides a way to achieve secure de-duplication (space-efficient secure outsourced storage), a goal currently targeted by numerous cloud-storage providers. We provide definitions both for privacy and for a form of integrity that we call tag consistency. Based on this foundation, we make both practical and theoretical contributions. On the practical side, we provide ROM security analyses of a natural family of MLE schemes that includes deployed schemes. On the theoretical side the challenge is standard model solutions, and we make connections with deterministic encryption, hash functions secure on correlated inputs and the sample-then-extract paradigm to deliver schemes under different assumptions and for different classes of message sources. Our work shows that MLE is a primitive of both practical.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

C.Security proofs for identity-based identification and signature schemes:

This paper provides either security proofs or attacks for a large number of identity-based identification and signature schemes defined either explicitly or implicitly in existing literature. Underlying these is a framework that on the one hand helps explain how these schemes are derived and on the other hand enables modular security analyses, thereby helping to understand, simplify, and unify previous work. We also analyze a generic folklore construction that in particular yields identity-based identification and signature schemes without random oracles.

D.A reverse de-duplication storage system optimized for reads to latest backups:

De-duplication is known to effectively eliminate duplicates, yet it introduces fragmentation that degrades read performance. We propose RevDedup, a de-duplication system that optimizes reads to the latest backups of virtual machine (VM) images using reverse de-duplication. In contrast with conventional de-duplication that removes duplicates from new data, RevDedup removes duplicates from old data, thereby shifting fragmentation to old data while keeping the layout of new data as sequential as possible. We evaluate our RevDedup prototype using a 12-week span of real-world VM image snapshots of 160 users. We show that RevDedup achieves high deduplication efficiency, high backup throughput, and high read throughput.

E.Secure de-duplication with efficient and reliable convergent key management:

Data de-duplication is a technique for eliminating duplicate copies of data, and has been widely used in cloud storage to reduce storage space and upload bandwidth. Promising as it is, an arising challenge is to perform secure de-duplication in cloud storage. Although convergent encryption has been extensively adopted for secure de-duplication, a critical issue of making convergent encryption practical is to efficiently and reliably manage a huge number of convergent keys. This paper makes the first attempt to formally address the problem of achieving efficient and reliable key management in secure de-duplication.We first introduce a baseline approach in which each user holds an independent master key for encrypting the convergent keys and outsourcing them to the cloud. However, such a baseline key management scheme generates an enormous number of keys with the increasing number of users and requires users to dedicatedly protect the master keys. To this end, we propose De-key, a new construction in which users do not need to manage any keys on their own but instead securely distribute the convergent key shares across multiple servers. Security analysis demonstrates that De-key is secure in terms of the definitions specified in the proposed security model. As a proof of concept, we implement De-key using the Ramp secret sharing scheme and demonstrate that De-key incurs limited overhead in realistic environments.

III.SYSTEM MODEL:

1)Hybrid Architecture for Secure De-duplication:

At a high level, our setting of interest is an enterprise network, consisting of a group of affiliated clients (for example, employees of a company) who will use the S-CSP and store data with de-duplication technique. The S-CSP performs de-duplication by checking if the contents of two files are the same and stores only one of them. Each privilege is represented in the form of a short message called token. Each file is associated with some file tokens, which denote the tag with specified privileges. Role of the private cloud server will be explained in the paper. block-level de-duplication can be easily deduced from file-level de-duplication. Specifically, to upload a file, a user first performs the file-level duplicate check. If the file is a duplicate, then all its blocks must be duplicates as well.

a)S-CSP: This is an entity that provides a data storage service in public cloud. The S-CSP provides the data outsourcing service and stores data on behalf of the users
b)Data Users: A user is an entity that wants to outsource data storage to the S-CSP and access the data later. In a storage system supporting de-duplication, the user only uploads unique data but does not upload any duplicate data to save the upload bandwidth, which may be owned by the same use or different users. Every single file is protected with the convergent encryption key and privilege keys to realize the authorized de-duplication with differential privileges.
c)Private Cloud: Compared with the traditional de-duplication architecture in cloud computing, this is a new entityintroduced for facilitating user's secure usage of cloud service. Private Keys are managed by private cloud in order to give them privileges as per their designation.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

2)Design Goals:

We have proposed a new de-duplication system for the following:

a)Differential Authorization: Each authorized user is able to get his/her individual token of his file to perform duplicate check based on his privileges. Under this assumption, any user cannot generate a token for duplicate check out of his privileges or without the aid from the private cloud server.
b)Authorized Duplicate Check: Authorized user is able to use his/her individual private keys to generate query for certain file and the privileges he/she owned with the help of private cloud, while the public cloud performs duplicate check directly and tells the user if there is any duplicate.
c)Unforgeability of file token/duplicate-check token: Unauthorized users without appropriate privileges or file should be prevented from getting or generating the file tokens for duplicate check of any file stored at the S-CSP. The duplicate check token of users should be issued from the private cloud server in our scheme.

PROBLEM STATEMENT:

•Data de-duplication systems, the private cloud is involved as a proxy to allow data owner/users to securely perform duplicate check with differential privileges.
•Such architecture is practical and has attracted much attention from researchers.
•The data owners only outsource their data storage by utilizing public cloud while the data operation is managed in private cloud.

Drawbacks:

•Traditional encryption, while providing data confidentiality, is incompatible with data de-duplication.
•Identical data copies of different users will lead to different ciphertexts, making de-duplication impossible.

PROBLEM DEFINITION:

In this paper, we enhance our system in security. Specifically, we present an advanced scheme to support stronger security by encrypting the file with differential privilege keys.

In this way, the users without corresponding privileges cannot perform the duplicate check. Furthermore, such unauthorized users cannot decrypt the cipher text even collude with the S-CSP. Security analysis demonstrates that our system is secure in terms of the definitions specified in the proposed security model.

Advantages:

» The user is only allowed to perform the duplicate check for files marked with the corresponding privileges.
» We present an advanced scheme to support stronger security by encrypting the file with differential privilege keys.
» Reduce the storage size of the tags for integrity check. To enhance the security of de-duplication and protect the data confidentiality.

IMPLEMENTATION:

1.Cloud Service Provider:

•In this module, we develop Cloud Service Provider module. This is an entity that provides a data storage service in public cloud.
•The S-CSP provides the data outsourcing service and stores data on behalf of the users.
•To reduce the storage cost, the S-CSP eliminates the storage of redundant data via de-duplication and keeps only unique data.
•In this paper, we assume that S-CSP is always online and has abundant storage capacity and computation power.

2.Data Users:

•A user is an entity that wants to outsource data storage to the S-CSP and access the data later.
•In a storage system supporting de-duplication, the user only uploads unique data but does not upload any duplicate data to save the upload bandwidth, which may be owned by the same user or different users.
•In the authorized de-duplication system, each user is issued a set of privileges in the setup of the system. Each file is protected with the convergent encryption key and privilege keys to realize the authorized de-duplication with differential privileges.

3.Private Cloud:

•Compared with the traditional de-duplication architecture in cloud computing, this is a new entity introduced for facilitating user's secure usage of cloud service.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

•Specifically, since the computing resources at data user/owner side are restricted and the public cloud is not fully trusted in practice, private cloud is able to provide data user/owner with an execution environment and infrastructure working as an interface between user and the public cloud.
•The private keys for the privileges are managed by the private cloud, who answers the file token requests from the users. The interface offered by the private cloud allows user to submit files and queries to be securely stored and computed respectively.

4.Secure De-duplication System:

•We consider several types of privacy we need protect, that is, i) Unforgeability of duplicate-check token: There are two types of adversaries, that is, external adversary and internal adversary.
•As shown below, the external adversary can be viewed as an internal adversary without any privilege.
•If a user has privilege p, it requires that the adversary cannot forge and output a valid duplicate token with any other privilege p′ on any file F, where p does not match p′. Furthermore, it also requires that if the adversary does not make a request of token with its own privilege from private cloud server, it cannot forge and output a valid duplicate token with p on any F that has been queried.

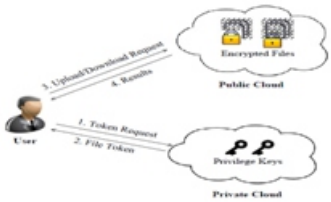

fig: Architecture of Authorized de-duplication

Our implementation of the Client provides the following function calls to support token generation and de-duplication along the file upload process.

• FileTag(File) - It computes SHA-1 hash of the File as File Tag;
• TokenReq(Tag, UserID) - It requests the Private Server for File Token generation with the File Tag and User ID;
• DupCheckReq(Token) - It requests the Storage Server for Duplicate Check of the File by sending the file token received from private server;
• ShareTokenReq(Tag, {Priv.}) - It requests the Private Server to generate the Share File Token with the File Tag and Target Sharing Privilege Set;
• FileEncrypt(File) - It encrypts the File with Convergent Encryption using 256-bit AES algorithm in cipher block chaining (CBC) mode, where the convergent key is from SHA-256 Hashing of the file; and
• FileUploadReq(FileID, File, Token) - It uploads the File Data to the Storage Server if the file is Unique and updates the File Token stored.

Our implementation of the Private Server includes corresponding request handlers for the token generation and maintains a key storage with Hash Map.

• TokenGen(Tag, UserID) - It loads the associated privilege keys of the user and generate the token with HMAC-SHA-1 algorithm; and
• ShareTokenGen(Tag, {Priv.}) - It generates the share token with the corresponding privilege keys of the sharing privilege set with HMAC-SHA-1 algorithm.

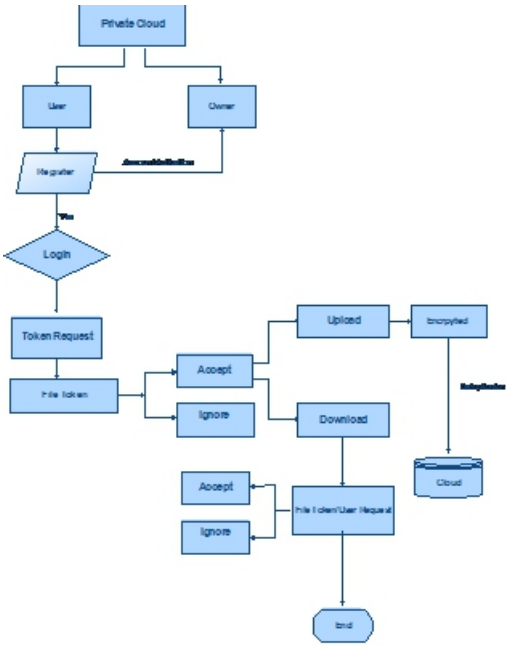

Fig:- flow

Our implementation of the Storage Server provides de-duplication and data storage with following handlers and maintains a map between existing files and associated token with Hash Map.

• Dup-Check(Token) - It searches the File to Token Map for Duplicate; and
• FileStore(FileID, File, Token) - It stores the File on Disk and updates the Mapping.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

CONCLUSION:

We additionally exhibited a few new de-duplication developments supporting approved copy weigh in half-breed cloud structural planning, where the copy check tokens of documents are produced by the private cloud server with private keys. Security examination exhibits that our plans are secure as far as insider and pariah assaults indicated in the proposed security model. As a proof of idea, we executed a model of our proposed approved copy check plan and behavior proving ground probes our model. We demonstrated that our approved copy check plan causes negligible overhead contrasted with united encryption and system exchange.

REFERENCES:

[1] OpenSSL Project. http://www.openssl.org/.

[2] P. Anderson and L. Zhang. Fast and secure laptop backups with encrypted de-duplication. In Proc. of USENIX LISA, 2010.

[3] M. Bellare, S. Keelveedhi, and T. Ristenpart. Dupless: Serveraided encryption for deduplicated storage. In USENIX Security Symposium, 2013.

[4] M. Bellare, S. Keelveedhi, and T. Ristenpart. Message-locked encryption and secure deduplication. In EUROCRYPT, pages 296– 312, 2013.

[5] M. Bellare, C. Namprempre, and G. Neven. Security proofs for identity-based identification and signature schemes. J. Cryptology, 22(1):1–61, 2009.

[6] M. Bellare and A. Palacio. Gq and schnorr identification schemes: Proofs of security against impersonation under active and concurrent attacks. In CRYPTO, pages 162–177, 2002.

[7] S. Bugiel, S. Nurnberger, A. Sadeghi, and T. Schneider. Twin clouds: An architecture for secure cloud computing. In Workshop on Cryptography and Security in Clouds (WCSC 2011), 2011.

[8] J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M. Theimer. Reclaiming space from duplicate files in a serverless distributed file system. In ICDCS, pages 617–624, 2002.

[9] D. Ferraiolo and R. Kuhn. Role-based access controls. In 15th NIST-NCSC National Computer Security Conf., 1992.

[10] GNU Libmicrohttpd. http://www.gnu.org/software/libmicrohttpd/.

[11] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg. Proofs of ownership in remote storage systems. In Y. Chen, G. Danezis, and V. Shmatikov, editors, ACM Conference on Computer and Communications Security, pages 491–500. ACM, 2011.

[12] 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.

[13] libcurl. http://curl.haxx.se/libcurl/.

[14] C. Ng and P. Lee. Revdedup: A reverse deduplication storage system optimized for reads to latest backups. In Proc. of APSYS, Apr 2013.