DEVELOPING AN AUTHORIZED DEDUPLICATION SYSTEM IN HYBRID CLOUD MODEL

MANDAVILLI GAYATHRI DEVI [#1] and KALADI GOVINDARAJU[*2]

[#] *PG Scholar, Kakinada Institute Of Engineering & Technology Department of Computer Science , JNTUK,A.P, India.*

[*] *Assistant Prof, Dept of CSE, Kakinada Institute of Engineering & Technology, JNTUK, A.P, INDIA.*

***Abstract*— In the recent years, the advancements in the cloud computing attracts several users. The cloud Service Provider should take responsibility of the data after its being uploaded by the cloud user. Several users can store the similar sorts of data, which incurs high storage space. In order to resolve this issue, data deduplication system is introduced. In this paper, we make an attempt to secure the data in the hybrid cloud model. To protect the data in the cloud, different privileges are provided for the users. We propose an improved secure authentication deduplication systems, which maintains the unique content of the data. The proposed scheme executes in four steps namely, Cloud service provider, Data users, Private cloud module and secure deduplication systems. The security is enhanced by two steps, Identity Provider and Indexing service. Experimental analysis and results shows the effectiveness of the proposed scheme.**

***Index Terms*— Cloud computing, Hybrid cloud, Authentication, Deduplication systems and file systems.**

I. INTRODUCTION

Due to the developments in cloud computing system, the data protection becomes a vital part of the system. Profound accessible storage and greatly parallel processing assets giving by the cloud administrations at low expenses. Most extreme measure of information put away in the cloud and shred by the clients with determined rights, which characterize as access privileges of the saved information. The critical challenge in the cloud storage services is the duplication. The duplication technique belongs to the class of data compression system. It has pulled in more consideration as of late. In the data storage, to decrease the information duplicates we go for duplication strategies. The strategy is utilized to enhance the usage of storage services and also to reduce the size of the data. Deduplication dispenses with repetitive information to decrease various information duplicates with the same content. Duplication just keeps one physical copy and eludes other repetitive information to that duplicate. Either the record level or block level, deduplication can occur.

Though, deduplication emits security and privacy concerns, yet the users are susceptible to the internal and external attacks. Customary encryption, while giving information classification, is inconsistent with information deduplication. In particular, customary encryption requires diverse clients to scramble their information with their own particular keys. Therefore, indistinguishable information duplicates of various clients will prompt distinctive ciphertexts, making deduplication incredible. Joined encryption has been proposed to implement data privacy while making deduplication in practical. It unscrambles information with secret key, which is acquired by registering the cryptographic hash estimation of the content. Clients hold the keys and send the ciphertext to the cloud after completing the key generation process. To avert unapproved access, a safe confirmation of proprietorship convention is likewise expected to give the verification that the client without a doubt claims the same document when a copy is found.

The rest of the paper is organized as follows: Section II describes the related work in the data deduplication systems. Section III discusses about the proposed methodology. It experiments and validation will discuss in Section IV. Atlast concludes in Section V.

II. LITERATURE SURVEY

There is a colossal measure of copy information or repetitive information, are available in storage systems. Excessively, it requires additional power supply and resource utilization. As the data scale increases, the maintenance and management of the data also increases. Thus, deduplication is introduced, to reduce the duplicate data. The semantic information de-duplication (SDD) is proposed, which makes utilization of the semantic data in the I/O way, of the documented records to coordinate the document into semantic parts (SP). The target of the SDD is to minimize the data duplication at the file level. This directly encloses SP into the disks with lot of fragments. Essential investigations have shown that SDD can assist to diminish the storage complexity when compared to existing techniques. With the appearance of distributed computing, secure information deduplication has pulled in much consideration, as of late, from the research community.

Yuan et al proposed a deduplication framework in the distributed storage to lessen the capacity size of the labels for data integrity. To upgrade the security of deduplication and ensure the information privacy, Bellare et al demonstrated to ensure the information privacy by changing the predictable message into unpredictable message. In their framework, another outsider called key server is acquainted with produce the document tag for duplicate check. *Stanek et al* exhibited a novel encryption plan that gives the vital security to prominent information and disagreeable information. For prominent information that are not especially delicate, the customary traditional encryption is performed. Along these lines, they accomplished better exchange between the proficiency and security of the outsourced information. *Li et al* tended to the key administration issue in block level deduplication by disseminating these keys over different servers subsequent to scramble the records.

"Differential authorized de-duplication check" cannot supported by the previous de-duplication systems. With the authorized de-duplication system, each user issued a set of the privileges during system initialization. To specify which type of user is allowed to perform the duplication check and access the files is decided by the uploading each file to the cloud and is also bounded by the set of privileges. The user have to take the file and the own privileges as inputs, to submit before of the user duplication check request for the same file. If only, copy of the file and matched privilege stored in cloud, then only the user gets the duplicate of the same file.

III. PROPOSED SECURE AUTHENTICATION DEDUPLICATION SCHEME

In this section, we explain about the improved secured authentication deduplication scheme. The system model contains entities such as Data user, Cloud Service provider, and Cloud server.

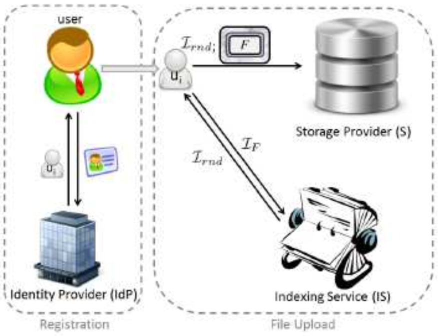


Fig.1. System Architecture

The proposed work segments into four modules:

a) Cloud service Provider: Let us assume that CSP is in online with the required storage and computational power. The task of CSP is to provide storage service in the public cloud. The data outsourcing service and saves the user's data on behalf of them. The target of CSP is to eliminate the duplicate data in order to reduce the storage cost.

b) Data user module: A user is an actor, who can upload, download, modify and access the data at anywhere in anytime. The user doesn't allow storing the content that is similar in nature. In case of authorized deduplication system, every user is defined by privileges, while uploading the data. Each file is protected by an encryption and privilege keys to differentiate the privileges.

c) Private Cloud Module: This is the new module introduced for providing secure usage of cloud service. In the private cloud, some services are being restricted by the community. It will act as an interface between user and cloud service provider. The files are accessed by the several numbers of tokens. This token will be used for further file access.

d) Secure Deduplication system: Though, we set different privacy level, there is a chance for attackers. The attacks may be of external and internal attacks. Some scenario, the external attackers can also consider as internal attackers. If the users possess a valid token with privilege p, then the file is enabled. It is executed in two components:

i) Identity Provider: The target of the IP is to prevent the Sybil attacks. The user is allowed to sign-in only once. It executes on users who possess encryption File E. This aim to create trust between the users and cloud service provider and also to maintain the confidentiality of the data.

ii) Indexing Service: It is the second Trusted Third Party. It ensures the leakage prevention of the unpopular files. The indexing service of the file is the determinant function i.e one-way function. In order to avoid the duplicate data, the indexing service is used. It also eliminates multiple upload of the data. The deduplication scheme executes before the indexing service, so as to secure the data.

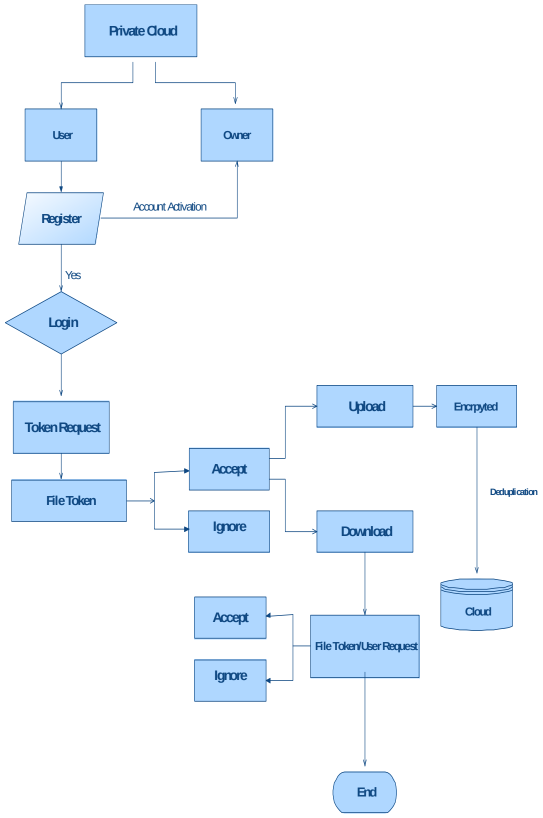


Fig.2. Proposed workflow

IV. EXPERIMENTAL RESULTS

In this section, we provide an experimental validation of proposed data deduplication scheme. Performance analysis is done in terms of storage space and computation.

A. Storage Space:

Consider a scenario of analyzing 857 files systems. The average number of files per file system is 225K and its size is 318K. By using secure Diffie Hellman key exchange system, the key size of each block is 265bits in our design. The storage space of metadata is calculated by four main data structures:

i) Linked list: It contain one node and multiple links for each block. Each file id and encrypted block keys hold 256 bits.

ii) Pointer Table: One record stores the pointer table. It contain block id and original block id is stored at CSP, 64 bits.

iii) Signature Table: It stores one record for each block which is non-duplicated. Each block contain block id, file id and signature bits.

iv) File table: It also use one record for storage. Each record contains file id, file name, user id and id of the first data block.

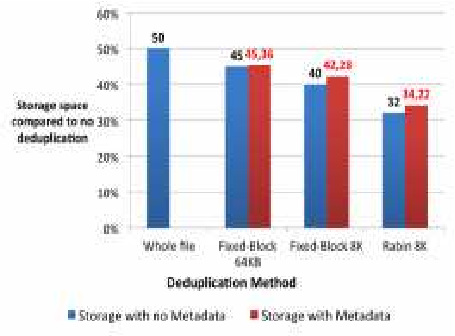


Fig.3. Storage analysis

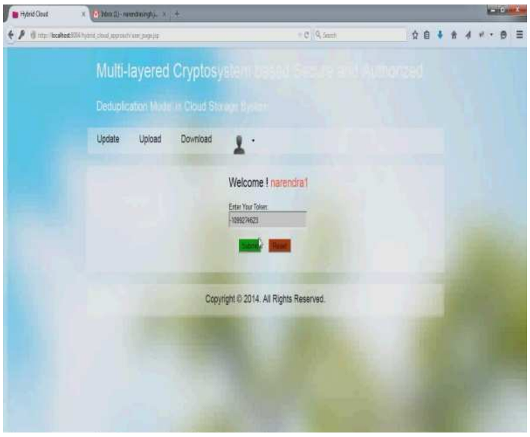
B. Computation analysis:

In the cloud storage, computational cost on storage and retrieval is the most analyzed performance validation. Let N be the number of blocks per file and M, the total number of blocks stored at the server.

	Storage	Retrieval
Encryption	O(N)	O(N)
Hash	O(N)	O(N)
Lookup in data structures	O(N log M)	O(N)

C. Security:

In the improved secure authentication deduplication scheme, the security is adopted in generating the dynamic tokens.


Hybrid Cloud
Inbox (1) - narendrasingh...
http://localhost:8084/hybrid_cloud_approach/user_page.jsp
Search
Update
Upload
Download
Submit
Reset

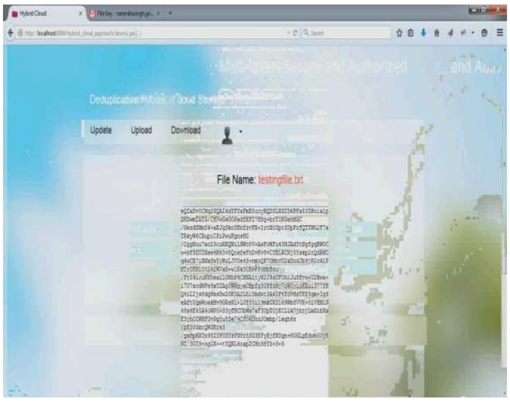
Multi-layered Cryptosystem based Secure and Authorized Deduplication Model in Cloud Storage System

Welcome ! narendra1

Enter Your Token: 1099274623

Copyright © 2014. All Rights Reserved.

Fig. 4. User entering the token mailed by the CSP.



File Name: testingfile.txt

Fig.5. File is decrypted only by the particular user.

V. CONCLUSION

In this paper, we study the issue of data confidentiality and security of deduplicated systems. We found that state-of-the–art suffers by external and internal attackers and computational overheads. We introduced improved Secure Authentication Deduplication systems that eliminates the duplicates files at both file level and block level file system. The security is enhanced by two steps: a) Identity Provider and b) Indexing service. The user who holds valid Identity Providers are allowed to access the file services. Experimental analysis is done in analyzing 857 files in file systems. The proposed scheme is validated in terms of storage, computational and security. Each user will be issued a secret key for their related users. The secret key further helps to generate tokens for their files. These file tokens should shared in order to upload, download or view the files of other users. Only the Identity Provider users are allowed to generate the token. Thus, in this way, a secure authentication deduplication system is achieved.

REFERENCES

[1] A Hybrid Cloud Approach for Secure Authorized Deduplication Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou (2014)

[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: Server aided 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.

[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 NISTNCSC National Computer Security Conf., 1992.

[10] GNULibmicrohttpd. 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.

[15] W. K. Ng, Y. Wen, and H. Zhu. Private data deduplication protocols in cloud storage. In S. Ossowski and P. Lecca, editors, Proceedings of the 27th Annual ACM Symposium on Applied Computing, pages 441–446. ACM, 2012.

AUTHOR PROFILE

MANDAVILLI GAYATHRI DEVI, is a student of Kakinada Institute Of Engineering & Technology affiliated to JNTUK, Kakinada pursuing M.Tech (Computer Science). Her Area of interest includes Cloud Computing and its objectives in all current trends and techniques in Computer Science.

KALADI GOVINDARAJU M.TECH is working as Assistant Professor, Department of Computer Science & Engineering, Kakinada Institute of Engineering & Technology, JNTUK, A.P, INDIA.