

Hybrid Cloud Approach for Secure Authorized Deduplication

**B.V. Somasekhar #1, P.Sandhya Lakshmi #2, G.Venkata Jyothsna #3,
S.Bharathi #4, D.Lokesh Reddy #5, G.Devarshi #6**

#1 Asst. Professor, Dept of Computer Science and Technology, Qis College of Engineering and Technology, Ongole

#2 Student, Dept of Computer Science and Technology, Qis College of Engineering and Technology, Ongole

#3 Student, Dept of Computer Science and Technology, Qis College of Engineering and Technology, Ongole

#4 Student, Dept of Computer Science and Technology, Qis College of Engineering and Technology, Ongole

#5 Student, Dept of Computer Science and Technology, Qis College of Engineering and Technology, Ongole

#6 Student, Dept of Computer Science and Technology, Qis College of Engineering and Technology, Ongole

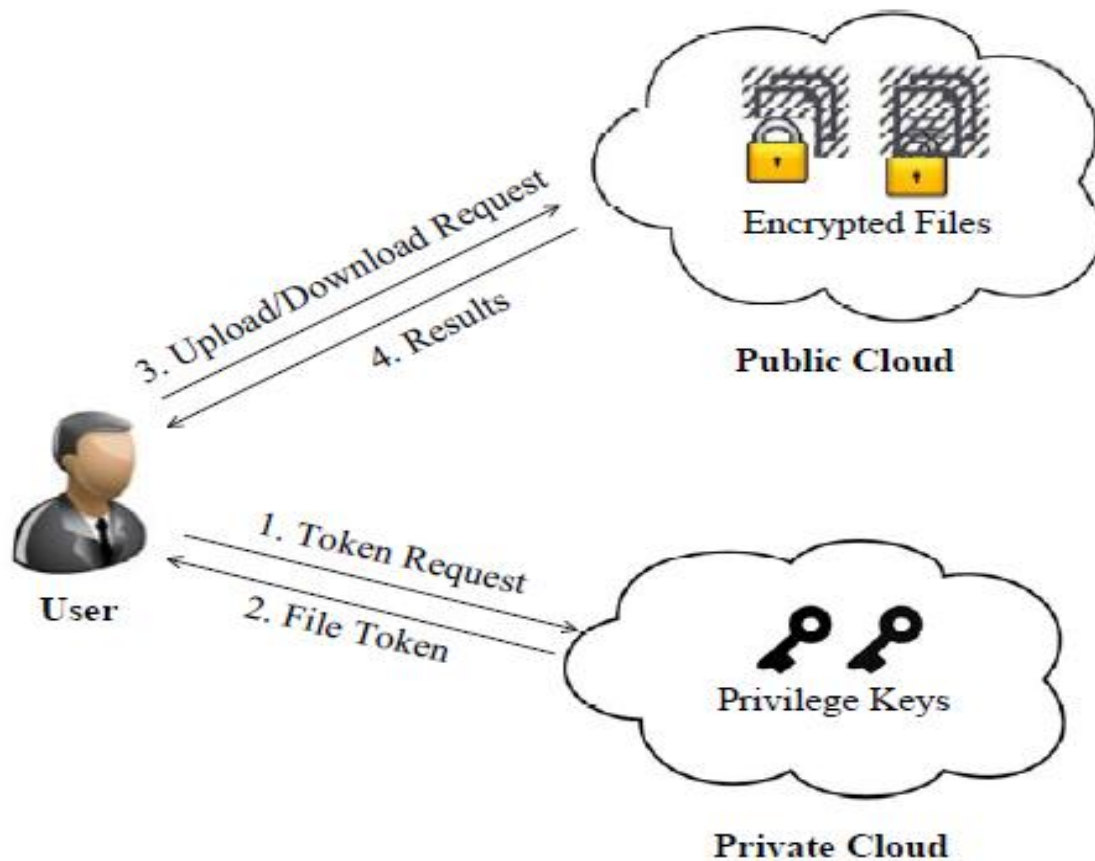
Abstract:

Today huge measure of information is being put away on the cloud. To make information the executives versatile, deduplication procedure is used. Data deduplication is one of significant information pressure strategies for forestalling various duplicates of same information, and has been broadly utilized in distributed storage to diminish the measure of extra room and to save transfer speed. To ensure the secrecy of delicate information the united encryption method has been utilized that is, encoding the information prior to reevaluating it. To help secure and approved deduplication, differential advantages of the clients are thought of. For better security this framework utilizes a mixture cloud engineering, which likewise bolsters the approved copy check.

Introduction:

One basic test that the distributed storage specialist organizations are confronting today is the administration of the always expanding volume of data. To make information the board adaptable in distributed computing, deduplication procedure is utilized. An information pressure strategy which is utilized to decrease the quantity of bytes sent is called Data Deduplication [17]. This procedure is utilized to improve capacity usage. Rather than keeping numerous duplicates of same information, deduplication keeps just a solitary duplicate of the information on the cloud, and a pointer is accommodated some other client attempting to transfer a similar record once more. There are two spots

where the Deduplication can be performed either at document level or the square level record level. The File level deduplication takes out numerous duplicates of same record, while the square level deduplication wipes out copy squares of information that happen in various documents. The Security and protection concerns emerge as the clients touchy information is helpless against assaults, despite the fact that it gives deduplication. To give security and protection in deduplication framework focalized encryption is utilized. Concurrent key is utilized to encode and unscramble the keys and code text is shipped off the cloud. The merged key is gotten by processing the hash estimation of the substance of the information of the record. Secure



verification of proprietorship convention is utilized to stay away from unapproved access. A client can download the scrambled document with the pointer from the worker, which must be decoded by the relating information proprietors with their merged keys. A half breed cloud engineering is utilized for giving security where in it comprises of a private cloud and a public cloud. Private cloud deals with all the encryption component and the public cloud stores all the scrambled documents. Information deduplication is one of significant information pressure strategies for killing copy duplicates of rehashing information, and has been generally utilized in distributed storage to lessen the measure of extra room and save data transmission. To ensure the classification of touchy

information while supporting deduplication, the concurrent encryption procedure has been proposed to scramble the information prior to rethinking. Security examination exhibits that our plan is secure as far as the definitions determined in the proposed security model. As a proof of idea, we execute a model of our proposed approved copy check plan and direct testbed tests utilizing our model. We show that our proposed approved copy check plot brings about negligible overhead contrasted with typical activities.

Related Work

Yuan et al. [24] proposed a deduplication framework in the distributed storage to

decrease the capacity size of the labels for uprightness check.

Bellare et al. [3] told the best way to secure the information classification by changing the anticipated message into capricious message. In their framework, another outsider called key worker is acquainted with create the record tag for copy check.

Stanek et al. [20] introduced a novel encryption plot that gives security to famous information and disagreeable information. The convention utilized scales well with huge individuals from records and the clients. Capacity enhancement strategies and start to finish encryption are the methods utilized. Record advances from one mode to different happens at the worker side if and just if a document becomes well known information.

Li et al. [12] proposed secure deduplication with productive and solid joined key administration. This paper recommends that the clients need not deal with the focalized keys, rather a calculation is proposed to deal with the merged keys.

Bellare et al. [4] proposed Message-bolted encryption and secure deduplication. Message-bolted encryption (MLE) is a cryptographic method, where a key under which encryption and unscrambling are performed is gotten from the message. MLE additionally permits to accomplish secure deduplication

P. Anderson et al. [2] proposed Fast and Secure Laptop Backups with Encrypted

Deduplication. Here a calculation and model programming where information is encoded autonomously without refuting the deduplication.

Halevi et al. [11] proposed a convention evidences of possession (PoW) for deduplication frameworks, the proposed framework builds up a security plot which can be utilized by the clients to demonstrate that the record is without a doubt claimed by him/her.

Bugiel et al. [7] proposed Twin mists: An engineering for secure distributed computing. This engineering permits secure rethinking of information. The calculations are part with the end goal that the believed cloud oversees security basic activities and ware cloud deals with the exhibition basic tasks.

Literature Survey:

Security proofs for identity-based identification and signature schemes by M. Bellare, C. Namprempe, and G. Neven 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.

Revdedup: A reverse deduplication storage system optimized for reads to latest backups, by C. Ng and P. Lee. Deduplication is known to effectively eliminate duplicates,

yet it introduces fragmentation that degrades read performance. RevDedup, a deduplication system that optimizes reads to the latest backups of virtual machine (VM) images using reverse deduplication is proposed. In contrast with conventional deduplication 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.

Secure deduplication with efficient and reliable convergent key management by P. Lee, and W. Lou. Data deduplication 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 deduplication in cloud storage. Although convergent encryption has been extensively adopted for secure deduplication, a critical issue of making convergent encryption practical is to efficiently and reliably manage a huge number of Research Article Volume 7 Issue No.5 International Journal of Engineering Science and Computing, May 2017 12767 <http://ijesc.org/> convergent keys. It makes the first attempt to formally address the problem of achieving efficient and reliable key management in secure deduplication. First it introduces a baseline approach in which each user holds an independent master key for encrypting the convergent keys and outsourcing them to the cloud.

System Study: Existing System

Data deduplication 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 deduplication.
- Identical data copies of different users will lead to different ciphertexts, making deduplication impossible

Proposed System

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.

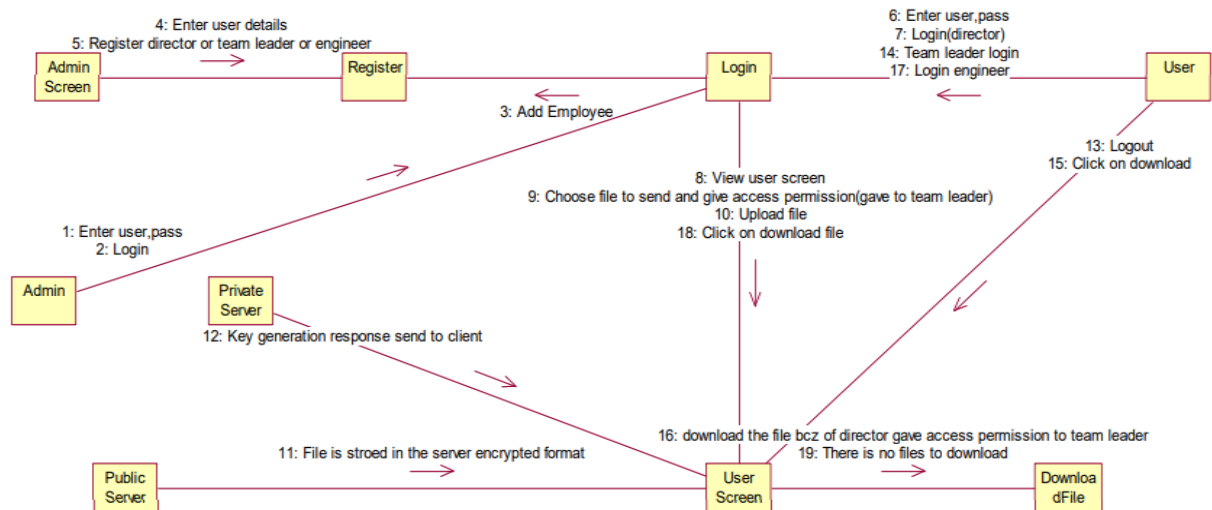
Advantages:

- 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 deduplication and protect the data confidentiality.

Implementation

- Cloud Service Provider
- Data Users Module
- Private Cloud Module



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 deduplication, 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 deduplication 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 deduplication with differential privileges.

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 deduplication and keeps only unique data. We assume that S-CSP is always online and has abundant storage capacity and computation power.

Data Users Module

Private Cloud Module

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

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/benefits 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.

Conclusion

The notion of authorized data deduplication was proposed to protect the data security by including differential privileges of users in the duplicate check. We also presented several new deduplication constructions supporting authorized duplicate check in hybrid cloud architecture, in which the duplicatecheck tokens of files are generated by the private cloud server with private keys. Security analysis demonstrates that our schemes are secure in terms of insider and outsider attacks specified in the proposed security model. As a proof of concept, we implemented a prototype of our proposed authorized duplicate check scheme and conduct test bed experiments on our prototype. We showed that our authorized duplicate check scheme incurs minimal overhead compared to convergent encryption and network transfer.

References

1. Open SSL 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: 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, pages 296–312, 2013.
5. M. Bellare, C. Namprempe, and G. Neven. Security proofs for Identity-based identification and signature schemes. J. Cryptology, 22(1):161, 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. 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 491500. 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 reversededuplication 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 441446. ACM, 2012.
16. R. D. Pietro and A. Sorniotti. Boosting efficiency and security in proof of ownership for deduplication. In H. Y. Youm and Y. Won, editors, ACM Symposium on Information, Computer and Communications Security, pages 8182. ACM, 2012.
17. S. Quinlan and S. Dorward. Venti: a new approach to archival storage. In Proc. USENIX FAST, Jan 2002.
18. A. Rahumed, H. C. H. Chen, Y. Tang, P. P. C. Lee, and J. C. S. Lui. A secure cloud backup system with assured deletion and version control. In 3rd International Workshop on Security in Cloud Computing, 2011.
19. R. S. Sandhu, E. J. Coyne, H. L. Feinstein, and C. E. Youman. Role-based access control models. IEEE Computer, 29:3847, Feb 1996.
20. J. Stanek, A. Sorniotti, E. Androulaki, and L. Kencl. A secure data deduplication scheme for cloud storage. In Technical Report, 2013.
21. M. W. Storer, K. Greenan, D. D. E. Long, and E. L. Miller. Secure data deduplication. In Proc. of StorageSS, 2008.
22. Z. Wilcox-O'Hearn and B. Warner. Tahoe: the least-authority filesystem. In Proc. of ACM StorageSS, 2008.
23. J. Xu, E.-C. Chang, and J. Zhou. Weak leakage-resilient client-side deduplication of encrypted data in cloud storage. In ASIACCS, pages 195206, 2013.
24. J. Yuan and S. Yu. Secure and constant cost public cloud storage auditing with deduplication. IACR Cryptology ePrint Archive, 2013:149, 2013.
25. K. Zhang, X. Zhou, Y. Chen, X. Wang, and Y. Ruan. Sedic: privacy aware data intensive computing on hybrid clouds. In Proceedings of the 18th ACM conference on Computer and communications security.

Authors Profile

B.V. Somasekhar, M.Tech., working as an Asst. Professor in the department of Computer Science & Engineering in QIS College of Engineering and Technology (Autonomous), Ongole, Andhra Pradesh, India.

P.Sandhya Lakshmi pursuing B Tech in Computer Science Engineering from QIS College of Engineering and Technology (Autonomous & NAAC 'A' Grade), Ponduru Road, Vengamukkalapalem, Ongole, Prakasam Dist, Affiliated to Jawaharlal Nehru Technological University, Kakinada.

G.Venkata Jyothsna pursuing B Tech in Computer Science Engineering from QIS College of Engineering and Technology (Autonomous & NAAC 'A' Grade), Ponduru Road, Vengamukkalapalem, Ongole, Prakasam Dist, Affiliated to Jawaharlal Nehru Technological University, Kakinada.

S.Bharathi pursuing B Tech in Computer Science Engineering from QIS College of Engineering and Technology (Autonomous & NAAC 'A' Grade), Ponduru Road, Vengamukkalapalem, Ongole, Prakasam Dist, Affiliated to Jawaharlal Nehru Technological University, Kakinada.

D.Lokesh Reddy pursuing B Tech in Computer Science Engineering from QIS College of Engineering and Technology (Autonomous & NAAC 'A' Grade), Ponduru Road, Vengamukkalapalem, Ongole, Prakasam Dist, Affiliated to

Jawaharlal Nehru Technological University, Kakinada.

G.Devarshi pursuing B Tech in Computer Science Engineering from QIS College of Engineering and Technology (Autonomous & NAAC 'A' Grade), Ponduru Road, Vengamukkalapalem, Ongole, Prakasam Dist, Affiliated to Jawaharlal Nehru Technological University, Kakinada.