

Detection of File Level and Block Level Deduplication and attacks in Cloud Computing Environment

Ms. Samita Mokal. Prof. Nilima D. Nikam. Prof. Vaishali Londhe

Shivajirao S. Jondhale College Of Engineering, Dombivli, India

patilsamita109@gmail.com

Yadavrao Tasgoankar Institute of Engineering and Technology, Bhivpuri Rd, India

nilu.nikam@gmail.com

Yadavrao Tasgoankar Institute of Engineering and Technology,
Bhivpuri Rd, India.

vaishali.londhe@tasgaonkartech.com

Abstract

In cloud computing, security and storage space management techniques are most important factors for improving the performance of cloud computing. Secure deduplication is a technique for eliminating duplicate copies of storage data, and provides security to them. To reduce storage space and upload bandwidth in cloud storage The basic idea in this paper is that we can eliminate duplicate copies of storage data and limit the damage of stolen data if we decrease the value of that stolen information to the attacker. Dekey 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 for insider attacker. User profiling and decoys, then, serve two purposes. First one is validating whether data access is authorized when abnormal information access is detected, and second one is that confusing the attacker with bogus information.

Keywords: Deduplication, proof of ownership, convergent encryption, key management, decoy technology.

1. INTRODUCTION

Cloud computing is model of the distribution of the information services in which the resources are the retrieved from the web through some of the interfaces and applications, instead forming direct connections to the server. Cloud storage systems provide the management of the ever increasing quantity of data by keeping in mind factors like reduce occupation storage space and the network bandwidth. Data deduplication also helps to improve the results in efficiency term and searches are quicker. Data deduplication may happen as file level deduplication or as block level data deduplication. Instead of maintaining numerous duplicate copies of file or the data with alike content, deduplication senses and remove the redundant data by keeping original physical copy. Data deduplication is a technique of eliminate duplicate copies of data, and it is used in cloud storage to reduce storage space and bandwidth. An arising challenge is to perform secure deduplication in cloud storage even if convergent encryption is extensively adopted for secure deduplication; a critical issue is that making of convergent encryption practical to manage a huge number of convergent keys efficiently and reliably.

The decoys, then, serve two purposes: First is that validating whether data access is authorized when abnormal information access is detected, and second one is that confusing the attacker with bogus information. Now this is

all about of outsider attacker protection while increase insider attacker with secure deduplication we are use convergent encryption [8] provides a viable option to enforce data confidentiality while realizing deduplication convergent key shares across multiple servers for insider attacker. As a proof of concept, we implement Dekey using the Ramp secret sharing scheme and demonstrate that Dekey incurs limited overhead in realistic environments.

2. LITERATURE REVIEW

Data deduplication is important for eliminating duplicate copies of data, and has been widely used in cloud storage to reduce storage space to users. They defined the notions used in based paper, review some secure primitives used secure deduplication. Symmetric Encryption, Convergent Encryption, Proofs of Ownership (PoWs), Ramp Secret Sharing, Secure Deduplication.

In 1997.M. Bellare, et.al explains that notion of security and scheme for Symmetric encryption in concentrate security framework. They give several differ notion of security and analyse the concrete complexity of reduction among them.

In 2002 John R. Douceur et al. Explain mechanism to reclaim space from this incidental duplication to make it available for controlled file replication. Their mechanism includes First one convergent encryption, which enables duplicate files to coalesced into the space of a single file, even if the files are encrypted with different users' keys, and second one SALAD, a Self- Arranging, Lossy, Associative Database for aggregating file content and location information in a decentralized, scalable, fault-tolerant manner.

In 2008 M.W.Storer[3] et.al developed two models for secure deduplicated storage authenticated and anonymous. These two designs demonstrate that security can be combined with deduplication in a way that provides a diverse range of security characteristics. In the models they present, security provided through the use of convergent encryption..

Halevi et al. [11] propose ‘ ‘ proofs of ownership’ ’ (PoW) for deduplication systems, such that a client can efficiently prove to the cloud storage server that he/she owns a file without uploading the file itself. Several PoW constructions based on the Merkle Hash Tree are proposed [11] to enable client-side deduplication, which include the bounded leakage setting. Pietro and Sorniotti [19] propose another efficient PoW scheme by choosing the projection of a file onto some randomly selected bit-positions as the file proof.

In 2012 M. Bellare et.al explain 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 deduplication (space-efficient secure outsourced storage), a goal currently targeted by numerous cloud-storage providers [7]. They provide definitions of privacy and integrity peculiar to this domain, exhibiting different trade between assumptions made and the message distributions for which security is proven. From our treatment MLE emerges as a primitive that combines practical impact with theoretical depth and challenges, making it well worthy of further study and a place in the cryptographic pantheon.

3. PROBLEM DEFINITION

We first develop outsourcing model used by Dekey. There are three entities, namely: the user, the cloud service provider Storage (CSP cloud service provider key-management (CSP KM), as categorized below. User: A user is an entity that outsource data storage to the S access the data later. To store bandwidth, the user only uploads data but does not upload any du which may be retained by the same user or different users.

CSP-S: The CSP-S provides the data outsourcing service and stores data on behalf of the users. To reduce the cost of storage the CSP-S removes the storage of redundant data via deduplication and keeps only unique data.

CSP-KM: A CSP-KM maintains convergent keys for users, and provides users with small storage and computation services to facilitate key management. In the fault tolerance of key management, we consider a quorum of KMs, each being a separate convergent key is distributed across multiple CSPs-KM using RSSS.

In this, we specify a data copy to be either a whole file or a smaller-size block, and this point to two types of deduplication: 1) file-level deduplication, which storage of any repeated files, and 2) block level deduplication, which fragment smaller fixed-size or variable-size blocks and delete the storage of any redundant bloc deploy our deduplication mechanism in both file level and block level. Specifically, to upload a file, a user first executes duplicate check. Dekey uses RSSS to provide a tunable key management mechanism to balance among confidentiality, ability, storage overhead, and performance Threat Model

and Security Goals Our threat model considers two types of attackers: 1) An outside attacker may obtain some knowledge (e.g., a hash value) of the data copy of interest via public channels. It plays a role of a user that interacts with the S-CSP. This kind of attacker includes the adversary who uses the S-CSP as an content distribution network; 2) An inside attacker is nest-but-curious, and it could refer to the S-CSP or any of the KM-CSPs. Its goal is to extract useful information of user data or convergent keys. We require the inside attacker to follow the protocol correctly. Here, we allow the collusion between the S-CSP and KM-CSPs.

4. PROPOSED SYSTEM

The basic idea in this paper is that we can eliminate duplicate copies of storage data and limit the damage of stolen data if we decrease the value of that stolen information to the attacker. This paper makes the first attempt to formally address the problem of achieving efficient and reliable key management in secure deduplication. We propose for providing security in insider attacker as well as outsider attacker and monitoring them we use for that Dekey, user behaviour profiling and Decoy Technology.

A new construction Dekey is proposed to provide efficient and reliable convergent key management through convergent key deduplication and secret sharing. Dekey supports both file-level and blocklevel deduplications.

We implement Dekey using the Ramp secret sharing scheme that enables the key management to adapt to different reliability and confidentiality levels. Our Evaluation demonstrates that Dekey incurs limited overhead in normal upload/download operations in realistic cloud environments.

We propose Dekey, an efficient and reliable convergent key management scheme for secure deduplication. Dekey applies deduplication among convergent keys and distributes convergent key shares across multiple key servers, while preserving semantic security of convergent keys and confidentiality of outsourced data.

5. SYSTEM ARCHITECTURE

5.1 Data Owner

In this module, the data owner uploads their data in the cloud server. For the security purpose the data owner encrypts the data file and then store in the cloud. The data owner can check the duplication of the file over Corresponding cloud server. The Data owner can have capable of manipulating the encrypted data file and the data owner can check the multiple cloud data as well as the duplication of the specific file. And also he can create remote user with respect to registered cloud servers. And also data owner has migrate to another cloud option, by this he can migrate files from one cloud server to another cloud server.

5.2 Cloud Server

The cloud service provider manages a cloud to provide data storage service. Data owners encrypt their data files and store them in the cloud for sharing with Remote

User. To access the shared data files, data consumers download encrypted data files of their interest from the cloud and then decrypt them.

5.3 Remote User

In this module, remote user logs in by using his user name and password. After he will request for secret key of required file from cloud servers, and get the secret key. After getting secret key he is trying to download file by entering file name and secret key from cloud server.

5.4 Attacker Module

In remote user module, while downloading time if remote user entered any wrong file name or secret key then cloud servers treat him as attacker and move his access permission to block/attacker list.

These shares will be distributed across multiple independent storage servers. Furthermore, to support deduplication, a short cryptographic hash value of the content will also be computed and sent to each storage server as the fingerprint of the fragment stored at each server. Only the data owner who first uploads the data is required to compute and distribute such secret shares, while all following users who own the same data.

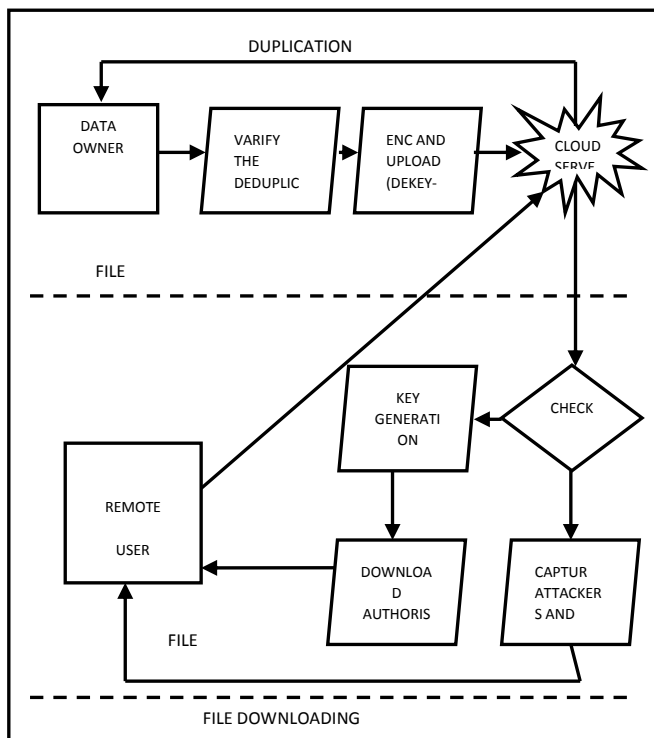


Figure 5.1 System Architecture

6. CONCLUSION

We propose Dekey, an efficient and reliable convergent key management scheme for secure deduplication. Dekey applies deduplication among convergent keys and distributes key shares across multiple key servers and provides confidentiality of outsourced data. Dekey implements small encoding/decoding overhead compared to the network transmission overhead in the regular upload/download operations. attacker module makes it more secure.

REFERENCES

- [1] Jin Li, Xiaofeng Chen, Mingqiang Li, Jingwei Li, Patrick P.C. Lee, and Wenjing louiee ,Secure Deduplication with Efficient and Reliable Convergent Key Management.IEEE transactions on parallel and distributed systems, vol. 25, no. 6, june 2014
- [2] M. Bellare, A. Desai, E. Jokipii, and P. Rogaway. A Concrete Security Treatment of Symmetric Encryption: Analysis of the DES Modes of Operation. Proceedings of the 38th Symposium on Foundations of Computer Science, IEEE, 1997.
- [3] Abdul Wahid Soomro, Nizamuddin, Arif Iqbal Umar, Noorul Amin.” Secured Symmetric Key Cryptographic Algorithm for Small Amount of Data” 3rd International Conference on Computer & Emerging Technologies (ICCET 2013)
- [4] M.W. Storer, K. Greenan, D.D.E. Long, and E.L. Miller, ,,,Secure Data Deduplication,”” in Proc. StorageSS, 2008, pp. 1-10.
- [5]W. J. Bolosky, J. R. Douceur, D. Ely, and M. Theimer, “ Feasibility of a Serverless istributed File System Deployed on an Existing Set of Desktop PCs” , SIGMETRICS 2000, ACM, 2000, pp.34-43.
- [6] A. Adya, W. J. Bolosky, M. Castro, R. Chaiken, G. Cermak, J. R. Douceur, J. Howell, J. R. Lorch, M. Theimer, and R. Wattenhofer.FARSITE: Federated, available, and reliable storage for an incompletely trusted environment. In Proceedings of the 5th Symposium on Operating Systems Design and Implementation (OSDI), Boston, MA, Dec.2002. USENIX.
- [7] A.D. Santis and B. Masucci, ,,,Multiple Ramp Schemes,”” IEEE Trans. Inf. Theory, vol. 45, no. 5, pp. 1720-1728, July 1999.
- [8] G.R. Blakley and C. Meadows, “Security of Ramp Schemes “, in Proc. Adv. CRYPTO, vol. 196, Lecture Notes in Computer Science,G.R. Blakley and D. Chaum, Eds., 1985, pp. 242-268.
- [9]M.O. Rabin, ,,,Efficient Dispersal of Information for Security, Load Balancing, Fault Tolerance,”” J. ACM, vol. 36, no. 2, pp. 335- 348, Apr. 1989.
- [10] A. Shamir, ,,,How to Share a Secret,”” Commun. ACM, vol. 22, no. 11, pp. 612-613, 1979.
- [11] NIST’ s Policy on Hash Functions, Sept. 2012. [Online]. Available: <http://csrc.nist.gov/groups/ST/hash/policy.html>.
- [12] AmazonCase Studies. [Online]. Available: <https://aws.amazon.com/solutions/case-studies/#backup>.
- [13] P. Anderson and L. Zhang, ‘ ‘ Fast and Secure Laptop Backups with Encrypted De-Duplication,’ ’ in Proc. USENIX LISA, 2010, pp. 1-8.
- [14] M. Bellare, S. Keelveedhi, and T. Ristenpart, ‘ ‘ Message-Locked Encryption and Secure Deduplication,’ ’ in Proc. IACR Cryptology ePrint Archive, 2012, pp. 296-3122012:631.
- [15] G.R. Blakley and C. Meadows, ‘ ‘ Security of Ramp Schemes,’ ’ in Proc. Adv. CRYPTO, vol. 196, Lecture Notes in Computer Science, G.R. Blakley and D. Chaum, Eds., 1985, pp. 242-268.
- [16] 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 Proc. ICDCS, 2002, pp. 617-624.

- [17] J. Gantz and D. Reinsel, The Digital Universe in 2020: Big Data, Bigger Digital Shadows, Biggest Growth in the Far East, Dec. 2012. [Online]. Available: <http://www.emc.com/collateral/analystreports/idc-the-digital-universe-in-2020.pdf>.
- [18] S. Halevi, D. Harnik, B. Pinkas, and A. Shulman-Peleg, ‘ ‘ Proofs of Ownership in Remote Storage Systems,’ ’ in Proc. ACM Conf. Comput. Commun. Security, Y. Chen, G. Danezis, and V. Shmatikov, Eds., 2011, pp. 491-500.
- [19] D. Harnik, B. Pinkas, and A. Shulman-Peleg, ‘ ‘ Side Channels in Cloud Services: Deduplication in Cloud Storage,’ ’ IEEE Security Privacy, vol. 8, no. 6, pp. 40-47, Nov./Dec. 2010.