Encryption Approaches for Secure Deduplication in Cloud Environment

Ketaki Ohale
Computer Department,
P. E. S. Modern
College of Engineering,
Shivajinagar, Pune-05

Aparna Junnarkar
Computer Department,
P. E. S. Modern
College of Engineering,
Shivajinagar, Pune-05

ABSTRACT
Digital data is growing voluminously due to the development in cloud services. This evolution is motivating users to outsource their data storage to the third party cloud providers. It is important for cloud storage services to remove redundancy among the data they store. Pirated duplication causes the revenue loss and violation of Intellectual Property Rights granted to the content owners. So, to protect cloud data from getting duplicated, deduplication techniques are widely used in the cloud based systems. Different approaches for deduplication of this cloud multimedia like audio, images, videos are compared and studied. These methods for secure deduplication improve the storage utilization and save network bandwidth.

General Terms
Cloud services, Deduplication.

Keywords
Convergent Encryption, Intellectual Property Rights (IPR), Message Locked Encryption (MLE), Public Key Cryptography, Ramp Secret Sharing Scheme (RSSS).

1. INTRODUCTION
Cloud storage is becoming increasingly popular amongst users due to its instant and easy access, wide variety of applications, scalability and reduced hardware costs. Also, it can act as a perfect file sharing tool for multimedia files. But along with this the process of duplication of the copyrighted multimedia like videos, images and music clips is also accelerated. Pirated duplication causes the revenue loss for its content holders and content creators. It also causes violation of the Intellectual Property Rights granted to the owners. It is a complex process to find the illegally made duplicate copies over the internet because of the sheer volume of the available multimedia content.

Multiple copies of the same data exist on these cloud servers. Due to this, the problem of increase in network and storage overhead caused at the servers arises. It is important for cloud storage services to remove redundancy among the data they store. Identifying common chunks of data both within and between the files regardless of the number of times they occur is called as deduplication. To minimize storage cost with overhead, de-duplication techniques are used in cloud infrastructures. Encryption is the most prominent approach when data security is taken in consideration.

Encryption strategies along with deduplication leads to secure data deduplication methods which are used by cloud based systems to manage their ever increasing volumes of data.

The paper is organized into four sections. Section 2 gives review of the Encryption Approaches for Secure Deduplication. Section 3 describes the performance parameters considered to compare these approaches and Finally, Section 4 summarizes and presents the conclusions.

2. RELATED WORK
The oldest approach to protect data and to achieve data security is encryption. In traditional cryptography, Symmetric key and public key encryption is used where Owner encrypts data with public key and client decrypts the same data with own private key.

Fingerprinting is the most popularly used variant of encryption. [10] Shows that randomly chosen irreducible polynomials are used to fingerprint bit-strings. This method is applied to produce a very simple real time string matching algorithm and procedure for securing files against unauthorized changes and deduplication. In traditional encryption, data privacy is maintained but same ciphertexts make deduplication almost impossible. Also it is not vulnerable to online or offline network attacks.

In cloud storage, Convergent encryption is a cryptosystem that provides data confidentiality by producing identical ciphertext from identical plaintext. Convergent encryption in various forms is used to eliminate duplicates in storage without the provider's access to the encryption keys. In Convergent Encryption as Message Locked Encryption, the key for encryption is itself derived from the message. Deterministic key encryption with hashing is used. Deduplication in server less distributed file system is achieved by using Convergent encryption. It is done by reclaiming spaces from duplicate files [7]. In this research, a scalable, server less, distributed file System Farasite is considered which is under development at Microsoft Research. SALAD, a Self-Arranging, Lossy, Associative Database for aggregating and analyzing file content information is proposed which is used for identification of identical encrypted files across a large number of machines in robust and decentralized manner.[2] Shows use of Block-Level Message-Locked Encryption (BL-MLE) for Secure Large File Deduplication. BL-MLE can achieve file-level and block-level deduplication, block key management, and proof of ownership simultaneously using a small set of metadata. Convergent encryption can be used in variety of schemes. [6] Formalized Message Locked Encryption (MLE) as a cryptographic primitive, where the key under which encryption and decryption are performed is itself derived from the message. MLE provides a way to achieve secure deduplication in space-efficient secure outsourced storage. Convergent encryption is vulnerable to brute force attacks. Also privacy and confidentiality is maintained with reduced reliability. But again same hash generation makes

deduplication impossible. Ramp Secret Sharing Schemes are another variant of Convergent encryption. In this scheme, original data copy is encrypted using convergent key and key is then encrypted using Master key. Identical ciphertexts are produced from identical plaintexts irrespective of encryption keys. Along with convergent key tag is derived which is used to identify duplicates. In [4] Convergent encryption with Ramp secret sharing scheme is used where Dekey is proposed due to which a small encoding and decoding overhead is achieved in the data upload/download operations. Encryption approaches for secure Deduplication in cloud environment posed using RSSS for the management of convergent keys. Further [1] shows improved reliability of the deduplications systems in distributed environment by performing file level and block level deduplications. In [8] and [9] Ramp scheme along with the Secret sharing schemes (SSSS) [11] is analyzed to distribute secret among the participants.

Different ramp schemes are reviewed to use in practical and fault tolerant environments for secure computation. A survey on secret sharing schemes is done in [3] where single secret sharing schemes are reviewed and RSSS is proved as threshold scheme used for fast computation and reduction of storage usage. Convergent Encryption using Ramp Secret Sharing Scheme is vulnerable to brute force as well as collusion attacks. They assure improved reliability, confidentiality and fault tolerance.

3. PERFORMANCE PARAMETERS

For secure deduplication in cloud based systems, encryption approaches can be compared on the basis of ideal block size, their Time complexities for storage and retrieval and encoding/decoding times.

3.1 Ideal block size

It is the optimal or absolute sequence of bits having maximum length suitable for the encryption.

3.2 Computational Complexity

It depends upon the time taken for encryption, hashing and lookup in the data structures. The time complexity of encryption approaches corresponds to the storage and retrieval of the data.

3.3 Encoding and Decoding time

Encoding time is the total time taken for hash generation and encryption. While decoding time is the time required for reconstructing the original data using the key. Usually encoding time is higher than the decoding time.

4. CONCLLUSION

Cloud computing is an emerging computing paradigm. Due to the advancements in the cloud infrastructure, there is need for multimedia content protection. Encryption and deduplication are the methods which are gaining importance for data backup and protection of cloud multimedia. Data encryption and deduplication are compatible with each other. But to use them together, encryption must take place first within the storage system and not at application or gateway layer. After the study and comparison of encryption approaches for deduplication of cloud multimedia it has been observed that convergent encryption using ramp secret sharing scheme takes lesser time for encoding/decoding the same amount of data. Thus, it is considered as an improvement in running time over traditional encryption. Deduplication systems using convergent encryption along with Ramp Secret sharing Scheme improves storage utilization, saves network bandwidth, storage cost and storage space efficiently.

5. ACKNOWLEDGMENT

Every orientation work has an imprint of many people and it becomes the duty of author to express the deep gratitude for the same. I feel immense pleasure to express deep sense of gratitude and indebtedness to my guide Prof. (Ms) Aparna Junnarkar, for constant encouragement and noble guidance.

I also express my sincere thanks to the Computer Department as well as Library of my college. Last but not the least; I am thankful to my friends and my parents whose best wishes are always with me.

6. REFERENCES

[1] Jin Li, Xiaofeng Chen, Xinyi Huang, Shaohua Tang and Yang Xiang ,"Secure Distributed Deduplication Systems with Improved Reliability", IEEE Transactions On Computers Volume, Year:2015.

[2] Rongmao Chen, Yi Mu, Guomin Yang, Fuchun Guo,"BL-MLE: Block-Level Message-Locked Encryption for Secure Large File Deduplication",IEEE Transactions on Information Forensics and Security, Volume: 10, Issue: 12, Dec 2015.

[3] Dharani P, Berlin M.A.,"Survey on Secret Sharing Scheme with Deduplication in Cloud Computing", IEEE 9th International Conference on Intelligent Systems and Control (ISCO),October 2015.

[4] 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, 2014.

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

[6] Mihir Bellare,Sriram Keelveedhi, Thomas Ristenpart, "Message-locked encryption and secure deduplication,"in EUROCRYPT ,2013.

[7] 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, 2002.

[8] A. D. Santis and B. Masucci, "Multiple ramp schemes," IEEE Transactions on Information Theory, vol. 45, no. 5, pp. 17201728, Jul. 1999.

[9] G. R. Blakley and C. Meadows, "Security of ramp schemes," in Advances in Cryptology: Proceedings of CRYPTO 84, ser. Lecture Notes in Computer Science, G. R. Blakley and D. Chaum, Eds. Springer-Verlag Berlin/Heidelberg,vol. 196, pp. 242268,1989.

[10] M. O. Rabin, "Fingerprinting by random polynomials," Center for Research in Computing Technology, Harvard University, Tech. Rep. Tech. Report TR-CSE-03-01,1981.

[11] A. Shamir, ”How to share a secret”, Commun. ACM, vol. 22, no. 11, pp. 612613, 1979.

[12] M. W. Storer, K. Greenan, D. D. E. Long, and E. L. Miller, ”Secure data deduplication”, in Proc. of StorageSS, 2008.

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