IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 3, March 2017

A Survey on Secure Authorized Deduplication in Hybrid Cloud

Ms. Chintu P Chacko

PG Scholar, Information Technology, TocH Institute of Science & Technology, Ernakulam, India

Abstract: Cloud storage has become very popular now a days because of its advantages like sharing of data among different geographical locations .But in most of the organizations, the cache can contain many duplicate copies of many pieces of same data. Many techniques being used for removing duplicate copies of repeating data, among these techniques one of the important data compression technique is data deduplication. Deduplication is widely used in cloud storage to save bandwidth and to reduce the amount of storage space. To protect the confidentiality of data along with deduplication convergent encryption technique is used before outsourcing. To effectively protect data security, this paper makes the first attempt to formally address the problem of authorized data deduplication.

Keywords: Deduplication, hybrid cloud, authorized duplicate check, convergent encryption, symmetric encryption.

I. INTRODUCTION

Cloud computing has recently appeared as a preferred business model for utility system. The formation of cloud is to supply computing resources as a utility or a service on demand to customers over the web. As cloud computing becomes widespread, an increasing amount of data is being stored in the cloud and shared by users with certain privileges. One of the critical challenges of the cloud storage service is the management of ever increasing volume of data. To make data management extensible in cloud computing, deduplication [1] has been a well-known technique used and has attracted more attention recently. Data deduplication is a data compression technique for removing duplicate copies of iterating data in storage. It is used to improve storage utilization and can to reduce the number of bytes that must be sent. Instead of keeping several data copies with the same content, deduplication eliminates redundant data by keeping only one physical copy. Although data deduplication brings endless benefits, security and privacy concerns arise due to attacks.

Traditional encryption, while providing data confidentiality, is conflicting with data deduplication. Traditional encryption requires alternate users to encrypt their information with reference to their own keys. Thus, identical data copies of users will lead to different cipher texts, causing deduplication impossible. Convergent encryption [2] has been proposed to insist upon data confidentiality while making deduplication feasible. Convergent encryption encrypts/decrypts a data copy with a convergent key, which is acquired by calculating the cryptographic hash value of the content of the data copy. To avoid unauthorized access, a secure proof of ownership protocol [3], [7] is also required to produce the proof that the user owns the same file when a duplicate is found.

In the model stated in this paper aims at efficiently solving the issue of deduplication with differential privileges in cloud computing. This model consist of a hybrid cloud architecture which includes both public cloud and private cloud. Unlike existing deduplication system private cloud is involved as a proxy to allow data owner to securely perform duplicate check with differential privileges. a new duplication system supporting differential duplicate check is proposed under this hybrid cloud architecture where the S-CSP [4] presides in the public cloud the user is only allowed to perform the duplicate check for files marked with corresponding privileges also present an advanced scheme to support strong security by encrypting the file with differential privilege case.

II. COMPARITIVE SURVEYING

This section focuses on some of the prior research work that addresses the security and storage issues in cloud computing technology.

Mihir Bellare, San Diego Sriram Keelveedhi, San Diego Thomas Ristenpart proposed DupLESS: Server-Aided Encryption for Deduplicated Storage **[6]** is used to provide secure deduplicated storage as well as storage resisting bruteforce attacks. Clients encrypt under message-based keys obtained from a key-server via an oblivious PRF protocol in dupless server. This allows clients to store encrypted data within an existing service and achieves strong confidentiality guarantees. This shows that encryption for deduplicated storage can reach desired performance and space savings near to that of using the storage service. In DupLESS, [6] customers encode under message-based keys acquired from a key-server by means of an absent PRF convention. It authorize customers to store disorganized information with a current administration, have the administration perform deduplication for their benefit, but then accomplishes solid privacy ensures.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 3, March 2017

Ng et al. [10] proposed RevDedup, a deduplication system that optimizes reads to latest VM image backups using an idea called reverse deduplication. In contrast with conventional deduplication that eliminates duplicates from new data, RevDedup [10] eliminates duplicates from old data, thereby shifting fragmentation to old data while keeping the layout of new data as sequential as possible. They evaluate their RevDedup prototype using micro benchmark and real-world workloads. For a 12-week span of real-world VM images from 160 users, RevDedup [10] achieves high deduplication efficiency and high backup and read throughput on the order of 1GB/s. RevDedup also incurs small metadata overhead in backup/read operations.

TABLE I COMPARITIVE STUUDY

Paper Title	Advantages	Disadvantages
Dupless: Server Aided Encryption for deduplicated storage	High performance	Failed to secure brute force attack
RevDedup: A Reverse Deduplication Storage System Optimized for Reads to Latest Backups	High deduplication productivity	Metadata overhead
CloudDedup: Secure Deduplication with Encrypted Data for Cloud Storage	Provides confidentiality	Does not impact the overall storage and computational cost
Proofs of ownership in remote storage systems	Time saving, rigorous security	Impossible to verify experimentally the assumption about the input distribution
Enhanced Dynamic whole file Deduplication for space optimization in private cloud storage backup	Reduce deduplication time, storage efficient, improve private cloud backup	Not sufficient to development of chunk and block level deduplication

J.Li, X.Chan, M.Li, P.Lee and W.Lou proposed [9] A Secure deduplication with efficient and reliable convergent key management, address the issue of accomplishing effective and dependable key administration in secure deduplication.

It represents a pattern approach in which every client holds an autonomous expert key for scrambling the focalized key and outsourcing them to cloud. This key administration plan produces a tremendous number of keys with the expanding number of clients and obliges clients to dedicatedly secure the expert keys. This mechanism reduces the storage space and bandwidth.

Pasqualo Puzio,Refik Molva and Malek Onen proposed [8]Cloud Dedup: Secure Deduplication with Encrypted Data for Cloud Storage, which assures block level deduplication and data confidentiality while coping with weaknesses raised by convergent encryption. [8] It preserves confidentiality and privacy even against potentially malicious cloud storage providers. It offers efficient key management solution through the metadata manager. It works transparently with existing cloud storage providers.

S.Halevi, D. Harnik,B.Pinkas and A. Shulman Peleg proposed Proofs of Ownership [3] in remote storage system, by which a client can prove to a server that it has a copy of a file without actually sending the file. This can be used to counter attacks on file deduplication system. It provides rigorous security [3]. Performance measurements indicate that the scheme incurs a small overhead compared to naive client side deduplication.

M.Shyamala Devi, V.Vimal Khanna,Naveen balaji proposed Enhanced Dynamic whole file Deduplication (DWFD) [5] for space optimization in private cloud storage backup, to optimize the private cloud storage backup inorder to provide high throughput to the user of the organization by increasing the deduplication efficiency. This schema [5] is not sufficient to the development of chunk level deduplication and block level deduplication.

III.CONCLUSION

The notion of authorized data deduplication was proposed in this paper to protect the data security by including differential privileges of users in the duplicate check. Also presented several new deduplication interpretations supporting authorized duplicate check in hybrid cloud architecture, in which duplicate check tokens of files are generated by the private cloud server with private keys. The authentication is based on the privileges provided for the user. This mechanism is highly secure against brute force attack and other passive attacks.

ACKNOWLEDGMENT

I gratefully thank the entire faculty of the Department of Information Technology, ToC H Institute of Science & Technology for their valuable support and encouragement in working on with this project work.

REFERENCES

[1] S. Quinlan and S. Dorward. Venti: a new approach to archival storage. In Proc. USENIX FAST, Jan 2002.

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 6, Issue 3, March 2017

[2] 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.
[3] 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
[4] Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou, "A Hybrid Cloud Approach for Secure Authorized Deduplication", IEEE Transactions on Parallel and Distributed Systems,Volume:PP,Issue:99,Date of Publication :18.April.2014
[5] M. Shyamala Devi, V.VimalKhanna,NaveenBalaji "Enhanced Dynamic Whole File De-Duplication(DWFD) for Space Optimization in Private Cloud Storage Backup",IACSIT, August,2014.
[6] M. Bellare, S. Keelveedhi, and T. Ristenpart. "Dupless: Server aided encryption for deduplicated storage". In USENIX SecuritySymposium, 2013.
[7] M. Bellare, S. Keelveedhi, and T. Ristenpart. Message-locked encryption and secure deduplication. In EUROCRYPT, pages 296–312, 2013.
[8] PasqualoPuzio, RefikMolva ,MelekOnen ,"CloudDedup: Secure Deduplication with Encrypted Data for Cloud Storage", SecludIT and EURECOM,France.
[9] 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.
[10] C. Ng and P. Lee. Revdedup: A reverse deduplication storage system optimized for reads to latest backups. In Proc. of APSYS, Apr 2013.

BIOGRAPHY

Chintu P Chacko received B. Tech degree in Information Technology from ICET, MG University, Kerala. Currently pursuing her Masters Degree in Network Computing from Toc H, APJ Abdul Kalam Technological University, Kerala, India.