International journal of basic and applied research
www.pragatipublication.com
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

SECURE CLOUD DATA DEDUPLICATION WITH EFFICIENT RE-ENCRYPTION

RAVI KIRAN KUMAR TERA 1, MEDI ARCHANA 2, NEERUDI BHOOMIKA 3, VISWAYAGNA JANARDHAN CHARY 4, ANTHAMOLLA RAMESH KUMAR 5, ASSISTANT PROFESSOR 1, UG SCHOLAR 2,3,4&5
DEPARTMENT OF CSE, MNR COLLEGE OF ENGINEERING AND TECHNOLOGY, MOHD.SHAPUR, TELANGANA 502285

ABSTRACT:-Data deduplication technique has been widely adopted by commercial cloud storage providers, which is both important and necessary in coping with the explosive growth of data. To further protect the security of users sensitive data in the outsourced storage mode, many secure data deduplication schemes have been designed and applied in various scenarios. Among these schemes, secure and efficient re- encryption for encrypted data deduplication attracted the attention of many scholars, and many solutions have been designed to support dynamic ownership management. In this paper, we focus on the re-encryption deduplication storage system and show that the recently designed lightweight rekeying- aware encrypted deduplication scheme (REED) is vulnerable to an attack which we

Keywords: Security, Cloud service Providers, deduplication, shared key

INTRODUCTION: Cloud technology is a trending buzzword among various environments. Cloud can access data and related files from any location and from any device at any time with an internet connection. As the working process is changing to flexible and remote working, it is essential to provide user-related data access to users, even when they are not at a workplace. Improved data security is another asset of cloud computing. With call it stub-reserved attack. Furthermore, we propose a secure data deduplication scheme with efficient re-encryption based on the convergent all-or-nothing transform (CAONT) and randomly sampled bits from the Bloom filter. Due to the intrinsic property of one-way hash function, our scheme can resist the stub-reserved attack and guarantee the data privacy of data owners' sensitive data. Moreover, instead of re-encrypting the entire package, data owners are only required to re-encrypt a small part of it through the CAONT, thereby effectively reducing the computation overhead of the system. Finally, security analysis and experimental results show that our scheme is secure and efficient in re- encryption. traditional data storage systems,

International journal of basic and applied research
www.pragatipublication.com
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

the data can be easily stolen or damaged. There can also be more chances for serious cyber attacks like viruses, malware, and hacking. Human errors and power outages can also affect data security. Use cloud computing, will get the advantages of improved data security. In the cloud, the data is protected in various ways such as anti-virus, encryption methods, and many more. Cloud technology is capable of providing better data storage, data security, collaboration. In this paper, we focus on the re-encryption deduplication storage system and show that the recently designed lightweight rekeying- aware encrypted deduplication scheme (REED) is vulnerable to an attack which we call it stub-reserved attack.

LITERATURE REVIEW

1. REKEYING FOR ENCRYPTED DEDUPLICATION STORAGE

Rekeying refers to an operation of replacing an existing key with a new key for encryption. It renews security protection, so as to protect against key compromise and enable dynamic access control in cryptographic storage. However, it is non-trivial to realize efficient rekeying in encrypted deduplication storage systems, which use deterministic content-derived encryption keys to allow deduplication on ciphertexts. We design and implement REED, a rekeying-aware encrypted deduplication storage system. REED builds on a deterministic version of all-or-nothing transform (AONT), such that it enables secure and lightweight rekeying, while preserving the deduplication capability. We propose two REED encryption schemes that trade between performance and security, and extend REED for dynamic access control. We implement a REED prototype with various performance optimization techniques. Our tracedriven testbed evaluation shows that our REED prototype maintains high performance and storage efficiency.

Drawbacks:

➢ Rekeying-aware encrypted deduplication storage system that aims for secure and lightweight rekeying, while preserving content similarity for deduplication.

International journal of basic and applied research
www.pragatipublication.com
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

➢ REED augments MLE with the idea of all-or nothing transforms (AONT) which transforms a secret into a package, such that the secret cannot be recovered without knowing the entire package.

➢ REED encrypts a small part of the package with a key that is subject to rekeying, while the remaining large part of the package is generated from a deterministic variant of AONT to preserve content similarity.

➢ Rekeying cannot be used to revoke users' access rights by re-encrypting ciphertexts (e.g., the genome data in the previous example) with new keys and making old keys inactive.

2. SECURE DISTRIBUTED DEDUPLICATION SYSTEMS WITH IMPROVED RELIABILITY

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. However, there is only one copy for each file stored in cloud even if such a file is owned by a huge number of users. As a result, deduplication system improves storage utilization while reducing reliability. Furthermore, the challenge of privacy for sensitive data also arises when they are outsourced by users to cloud. Aiming to address the above security challenges, this paper makes the first attempt to formalize the notion of distributed reliable deduplication system. We propose new distributed deduplication systems with higher reliability in which the data chunks are distributed across multiple cloud servers. The security requirements of data confidentiality and tag consistency are also achieved by introducing a deterministic secret sharing scheme in distributed storage systems, instead of using convergent encryption as in previous deduplication systems. Security analysis demonstrates that our deduplication systems are secure in terms of the definitions specified in the proposed security model. As a proof of concept, we implement the proposed systems and demonstrate that the incurred overhead is very limited in realistic environments.

Drawbacks:

International journal of basic and applied research
www.pragatipublication.com
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

• The distributed deduplication systems' is to not-reliably store data in the cloud while achieving confidentiality and integrity.

• It is to enable deduplication and distributed storage of the data across multiple storage servers not efficiently.

• Encrypting the data to keep the confidentiality of the data, our new constructions not utilize the secret splitting technique to split data into shards.

EXISTING SYSTEM: A number of deduplication systems have been proposed based on various deduplication strategies such as client-side or server-side deduplications, file-level or block-level deduplications. Bellare et al formalized this primitive as message-locked encryption, and explored its application in space efficient secure outsourced storage. There are also several implementations of convergent implementations of different convergent encryption variants for secure deduplication. Li addressed the key- management issue in block-level deduplication by distributing these keys across multiple servers after encrypting the files. showed how to protect data confidentiality by transforming the predictable message into a unpredictable message. Data reliability is actually a very critical issue in a deduplication storage system because there is only one copy for each file stored in the server shared by all the owners. Most of the previous deduplication systems have only been considered in a single-server setting. The traditional deduplication methods cannot be directly extended and applied in distributed and multi-server systems.

PROPOSED SYSTEM: In this paper, we show how to design secure deduplication systems with higher reliability in cloud computing. We introduce the distributed cloud storage servers into deduplication systems to provide better fault tolerance. To further protect data confidentiality, the secret sharing technique is utilized, which is also compatible with the distributed storage systems. In more details, a file is first split and encoded into fragments by using the technique of secret sharing, instead of encryption mechanisms. 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

International journal of basic and applied research
www.pragatipublication.com
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

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 copy do not need to compute and store these shares any more. To recover data copies, users must access a minimum number of storage servers through authentication and obtain the secret shares to reconstruct the data. In other words, the secret shares of data will only be accessible by the authorized users who own the corresponding data copy. Four new secure deduplication systems are proposed to provide efficient deduplication with high reliability for file-level and block- level deduplication, respectively. The secret splitting technique, instead of traditional encryption methods, is utilized to protect data confidentiality. Specifically, data are split into fragments by using secure secret sharing schemes and stored at different servers.

ADVANTAGES OF PROPOSED SYSTEM:

❖ Distinguishing feature of our proposal is that data integrity, including tag consistency, can be achieved.

❖ To our knowledge, no existing work on secure deduplication can properly address the reliability and tag consistency problem in distributed storage systems.

❖ Our proposed constructions support both file-level and block-level deduplications.

❖ Security analysis demonstrates that the proposed deduplication systems are secure in terms of the definitions specified in the proposed security model. In more details, confidentiality, reliability and integrity can be achieved in our proposed system. Two kinds of collusion attacks are considered in our solutions. These are the collusion attack on the data and the collusion attack against servers. In particular, the data remains secure even if the adversary controls a limited number of storage servers.

❖ We implement our deduplication systems using the Ramp secret sharing scheme that enables high reliability and confidentiality levels. Our evaluation results demonstrate that the new proposed constructions are efficient and the redundancies are optimized and comparable with the other storage system supporting the same level of reliability.

International journal of basic and applied research
www.pragatipublication.com
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

MODULES DESCRIPTION: System Model In this first module, we develop two entities: User and Secure-Cloud Service Provide.

User: The 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. Furthermore, the fault tolerance is required by users in the system to provide higher reliability.

S-CSP: The S-CSP is an entity that provides the outsourcing data storage service for the users. In the deduplication system, when users own and store the same content, the S-CSP will only store a single copy of these files and retain only unique data. A deduplication technique, on the other hand, can reduce the storage cost at the server side and save the upload bandwidth at the user side. For fault tolerance and confidentiality of data storage, we consider a quorum of S- CSPs, each being an independent entity. The user data is distributed across multiple S-CSPs.

DATA DEDUPLICATION Data Deduplication involves finding and removing of duplicate datas without considering its fidelity.

- Here the goal is to store more datas with less bandwidth.
- Files are uploaded to the CSP and only the Dataowners can view and download it.
- The Security requirements is also achieved by Secret Sharing Scheme.
- Secret Sharing Scheme uses two algorithms, share and recover.
- Datas are uploaded both file and block level and the finding duplication is also in the same process.
- This is made possible by finding duplicate chunks and maintaining a single copy of chunks.

FILE LEVEL DEDUPLICATION SYSTEMS

- To support efficient duplicate check, tags for each file will be computed and are sent to S-CSPs.
- To upload a file F , the user interacts with SCSPs to perform the deduplication.

International journal of basic and applied research
www.pragatipublication.com
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

- More precisely, the user firstly computes and sends the file tag ϕF = TagGen(F) to S-CSPs for the file duplicate check.
- If a duplicate is found the user computesand sendsit to a server via a secure channel.
- Otherwise if no duplicate is found the process continues,i.e secret sharing scheme runs and the user will upload a file to CSP.
- To download a file the user will use the secret shares and download it from the SCSP's
- This approach provides fault tolerance and allows the user to remain accessible even if any limited subsets of storage servers fail

CONCLUSION We proposed the distributed deduplication systems to improve the reliability of data while achieving the confidentiality of the users' outsourced data without an encryption mechanism. Four constructions were proposed to support file- level deduplication. The security of tag consistency and integrity were achieved.

REFERENCES

[1] X. Chen, J. Li, J. Weng, J. Ma, and W. Lou, "Verifiable computation over large database with incremental updates," IEEE Trans. Computers, vol. 65, no. 10, pp. 3184–3195, 2016.

[2] M. Gerla, J. Weng, and G. Pau, "Pics-on-wheels: Photo surveillance in the vehicular cloud," International Conference on Computing, Networking and Communications, pp. 1123–1127, 2013.

[3] X. Chen, J. Li, J. Ma, Q. Tang, and W. Lou, "New algorithms for secure outsourcing of modular exponentiations," IEEE Trans. Parallel Distrib. Syst., vol. 25, no. 9, pp. 2386–2396, 2014.

[4] H. Yuan, X. Chen, T. Jiang, X. Zhang, Z. Yan, and Y. Xiang, "Dedupdum: Secure and scalable data deduplication with dynamic user management," Inf. Sci., vol. 456, pp. 159–173, 2018.

[5] H. Huang, X. Chen, Q. Wu, X. Huang, and J. Shen, "Bitcoinbased fair payments for outsourcing computations of fog devices," Future Generation Comp. Syst., vol. 78, pp. 850–858, 2018.

International journal of basic and applied research
www.pragatipublication.com
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

[6] IDC. (2014) The digital universe of opportunities : Rich data and the increasing value of the internet of things. [Online]. Available:https://www.emc.com/leadership/digitaluniverse/2014iview/index.htm