ISSN: 2455 - 9172

EIJO Journal of Engineering, Technology And Innovative Research (EIJO–JETIR)

Einstein International Journal Organization (EIJO)

Available Online at: www.eijo.in

Volume – 1, Issue – 1, March – April 2016, Page No. : 54 - 57

Improved Reliability for Secure Distributed Deduplication System

B. Priyanka[1], K. Sasikala[2], V. M. Suresh[3]

[1, 2, 3] Department of Information Technology

E.G.S Pillay Engineering College, Nagapattinam, Tamil Nadu, India

E-Mail: sasi.kala94@gmail.com, solai1977@gmail.com

ABSTRACT

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.

Keywords: Secure, Deduplication, System, eliminating, Cloud, Privacy.

1. INTRODUCTION

By the unpredictable development of digital data, deduplication techniques are broadly engaged to backup data and decrease network and storage transparency by notice and eradicate redundancy among data. As an alternative of maintaining multiple data copies with the same content, deduplication reducing redundant data by maintaining only single copy and referring other redundant data to that copy. Deduplication has inward much concentration from both academic world and industry since it can really recover storage utilization and keep storage space, particularly for the applications with high deduplication ratio such as archival storage systems.

2. Existing System:

The various kinds of data for each user stored in the cloud and the demand of long term continuous assurance of their data safety, the problem of verifying correctness of data storage in the cloud becomes even more challenging. Cloud Computing is not just a third party data warehouse. The data stored in the cloud may be frequently updated by the users, including insertion, deletion, modification, appending, reordering, etc.

One critical challenge of today's cloud storage services is the management of the ever-increasing volume of data. According to the analysis report of IDC, the volume of data in the wild is expected to reach 40 trillion gigabytes in 2020. The baseline approach suffers two critical deployment issues. First, it is inefficient, as it will generate an enormous number of keys with the increasing number of users. Specifically, each user must associate an encrypted convergent key with each block of its outsourced encrypted data copies, so as to later restore the data copies.

Disadvantages

1. Data reliability is actually a very critical issue in a deduplication storage system.

2. Deduplication methods cannot be directly extended and applied in distributed and multi-server systems.

3. Proposed System:

A Dekey, a 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. Dekey using the Ramp secret sharing scheme and demonstrate that Dekey incurs limited overhead in realistic environments Dekey, which provides efficiency and reliability guarantees for convergent key management on both user and cloud storage sides. 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 Deduplication. Security analysis demonstrates that Dekey is secure in terms of the definitions specified in the proposed security model. Dekey using the 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.

Advantages

1. The detection of masquerade activity.
2. The deterrence affect which, although hard to measure, in preventing masquerade activity by risk-averse attackers.

4. Modules

4.1. Secure Deduplication:

Data deduplication is a specialized data compression technique for eliminating duplicate copies of repeating data. Related and somewhat synonymous terms are intelligent (data) compression and single-instance (data) storage. This technique is used to improve storage utilization and can also be applied to network data transfers to reduce the number of bytes that must be sent. In the deduplication process, unique chunks of data, or byte patterns, are identified and stored during a process of analysis. As the analysis continues, other chunks are compared to the stored copy and whenever a match occurs, the redundant chunk is replaced with a small reference that points to the stored chunk. Given that the same byte pattern may occur dozens, hundreds, or even thousands of times (the match frequency is dependent on the chunk size), the amount of data that must be stored or transferred can be greatly reduced.

4.2. User behavior profiling:

The data access in the cloud and detect abnormal data access patterns User profiling is a well known Technique that can be applied here to model how, when, and how much a user accesses their information in the Cloud. Such 'normal user' behavior can be continuously checked to determine whether abnormal access to a user's information is occurring. This method of behavior-based security is commonly used in fraud detection applications. Such profiles would naturally include volumetric information, how many documents are typically read and how often and monitor for abnormal search behaviors that exhibit deviations from the user baseline the correlation of search behavior anomaly detection with trap-based decoy files should provide stronger evidence of malfeasance, and therefore improve a detector's accuracy.

4.3. Decoy Documents:

A different approach for securing data in the cloud using offensive decoy technology and monitor data access in the cloud and detect abnormal data access patterns. After that launch a disinformation attack by returning large amounts of decoy information to the attacker. This protects against the misuse of the user's real data. This technology to launch disinformation attacks against malicious insiders, preventing them from distinguishing the real sensitive customer data from fake worthless data the decoys, then, serve two purposes: Validating whether data access is authorized when abnormal information access is detected and confusing the attacker with bogus information.

5. Literature Review

Reclaiming Space from Duplicate Files in a Server less Distributed File System

The Far site distributed file system provides availability by replicating each file onto multiple desktop computers. Since this replication consumes significant storage space, it is important to reclaim used space where possible. Measurement of over 500 desktop file systems shows that nearly half of all consumed space is occupied by duplicate files. We present a mechanism to reclaim space from this incidental duplication to make it available for controlled file replication. Our mechanism includes 1) 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 2) SALAD, a Self Arranging, Lossy, Associative Database for aggregating file content and location information in a decentralized, scalable, fault-tolerant manner. Large-scale simulation experiments show that the duplicate-file coalescing system is scalable, highly effective, and fault-tolerant.

Dupless: Server-aided encryption for deduplicated storage

Cloud storage service providers such as Drop box, Mozy, and others perform deduplication to save space by only storing one copy of each file uploaded. Should clients conventionally encrypt their files, however, savings are lost. Message-locked encryption (the most prominent manifestation of which is convergent encryption) resolves this tension. However it is inherently subject to brute-force attacks that can recover files falling into a known set. We propose an architecture that provides secure deduplicated storage resisting brute-force attacks, and realize it in a system called DupLESS. In DupLESS, clients encrypt under message-based keys obtained from a key-server via an oblivious PRF protocol. It enables clients to store encrypted data with an existing service, have the service perform deduplication on their behalf, and yet

achieves strong confidentiality guarantees. We show that encryption for deduplicated storage can achieve performance and space savings close to that of using the storage service with plaintext data. Providers of cloud-based storage such as Drop box, Google Drive, and Mozy can save on storage costs via deduplication: should two clients upload the same file, the service detects this and stores only a single copy.

Message-locked encryption and secure deduplication

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. We provide definitions both for privacy and for a form of integrity that we call tag consistency. Based on this foundation, we make both practical and theoretical contributions. On the practical side, we provide ROM security analyses of a natural family of MLE schemes that includes deployed schemes. On the theoretical side the challenge is standard model solutions, and we make connections with deterministic encryption, hash functions secure on correlated inputs and the sample-then-extract paradigm to deliver schemes under different assumptions and for different classes of message sources. Our work shows that MLE is a primitive of both practical and theoretical interest.

Multiple ramp schemes

The problem of establishing bounds on the size of the shares to be given to participants in secret sharing schemes is one of the basic problems in the area and has received considerable attention by several researchers. The practical relevance of this issue is based on the following observations: Firstly, the security of any system tends to degrade as the amount of information that must be kept secret, i.e., the shares of the participants, increases. Secondly, if the shares given to participants are too long, the memory requirements for the participants will be too severe and, at the same time, the shares distribution algorithms will become inecient. Therefore, it is important to derive significant upper and lower bounds on the information distributed to participants. The problem of estimating the amount of random bits necessary to set up the schemes has also received considerable attention. This is due to the fact that the amount of randomness needed by an algorithm is to be considered a computational resource, analogously to the amount of time and space needed.

Secure deduplication with efficient and reliable convergent key management

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 deduplication has been a well-known technique. For that purpose convergent encryption has been extensively adopt for secure deduplication, critical issue of making convergent encryption practical is to efficiently and reliably manage a huge number of convergent keys. 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 first introduce a baseline approach in which each user holds an independent master key for encrypting the convergent keys and outsourcing them. However, such a baseline key management scheme generates an enormous number of keys with the increasing number of users and requires users to dedicatedly protect the master keys. To this end, we propose Dekey, User Behavior Profiling and Decoys technology. 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.

6. System Architecture

By the unpredictable development of digital data, deduplication techniques are broadly engaged to backup data and decrease network and storage transparency by notice and eradicate redundancy among data. As an alternative of maintaining multiple data copies with the same content, deduplication reducing redundant data by maintaining only single copy and referring other redundant data to that copy. Deduplication has inward much concentration from both academic world and industry since it can really recover storage utilization and keep storage space, particularly for the applications with high deduplication ratio such as archival storage systems.

7. Conclusion

The proposed distributed deduplication systems are to increase the consistency of data however attaining the privacy of the user's outsourced data without an encryption appliance. The security of tag consistency and integrity were attained. The implementation of deduplication systems using the Ramp secret sharing scheme here gives the demonstration that it acquires small encoding/decoding overhead compared to the network transmission overhead in regular download /upload operations.

8. Future Enhancement

For future work, how to prevent a duplicate faking or maliciously-generated cipher text replacement attack. A security notion of tag consistency has been formalized for this kind of attack. In a deduplication storage system with tag consistency, it requires that no adversary is able to obtain the same tag from a pair of different messages with a non-negligible probability. This provides security guarantees against the duplicate faking attacks in which a message can be undetectably replaced by a fake one. In the previous related work on reliable deduplication over encrypted data, the tag consistency cannot be achieved as the tag is computed by the data owner from underlying data files, which cannot be verified by the storage server. As a result, if the data owner replaces and uploads another file that is different from the file corresponding to the tag, the following users who perform the duplicate check cannot detect this duplicate faking attack and extract the exact files they want. To solve this security weakness, suggested to compute the tag directly from the cipher text by using a hash function. This solution obviously prevents the cipher text replacement attack because the cloud storage server is able to compute the tag by itself. However, such a method is unsuitable for the distributed storage system to realize the tag consistency. The challenge is that traditional secret sharing schemes are not deterministic.

9. References

[1]. Anderson,P., Zhang, L (2010)" Fast And Secure Laptop Backup With Encrypted De-duplication" In Proceeding Of USENIX LISA.

[2]. Attenise, C., Burns, R., Curtmola, R., Herring, J., Kissner, L., Peterson, Z., song, D(2007)" Provable Data Possession At Untrusted Stores" In Proceeding Of The 14th ACM Conference On Computer And Communication security,pp.598-609.

[3]. Bellare, M., Keelveedhi, S., and Ristenpart, T (2013) " Dupless: Server- Aided Encryption For Deduplicated Storage" In Proceeding Of USENIX Security Symposium.

[4].Halevi,S., Harnik,D., Pinkas,B., and Shulman-Peleg,A (2011)" Proofs Of Ownership In Remote Storage systems"In ACM Conference On computer And Communications security,pp.491-500.

[5]. Li,j., Chen,X., Li,M., Li,J., Lee,P., Lou,W.,(2014)"Secure Deduplication With Efficient And Reliable Convergent Key Management "In IEEE Transaction On Parallel And Distributed Systems,pp.1615-1625.

[6]. Li,M.,Qin,C.,Lee,P,P,C., and Li,J., (2014)"Convergent Dispersal: Toward Storage-Efficient In a Cloud Of Clouds" In the 6th USENIX Workshop on Hot Topics In Storage And File Systems.

[7]. Ng, W,k., Wen,Y.,Zha,H. (2012)" Private Data Deduplication Protocols In Cloud storage" In Proceedings Of The 27th Annual ACM Symposium On Applied Computing,pp.441-446.

[8]. Plank,J,S., Simmerman,S., Schuman,C,D (2008) " Jerasure: A Library in C/C++ Facilitating Erasure Coding For Storage Applications".

[9]. Plank,J,S.,Xu,L.,(2006)"Optimizing Cauchy Reed-Solomon Codes For Fault-Tolerant Network Storage Applications In 5th IEEE International Symposium On Network Computing Application.

[10]. Rahumed,A., Chen,H.C.H., Tang,Y., Lee,P.P.C., and Lui,J.C.S.(2011)"A Secure Cloud Backup System With Assured Deletion".