International Journal of Applied Research 2015; 1(9): 201-203

International Journal of Applied Research

ISSN Print: 2394-7500
ISSN Online: 2394-5869
Impact Factor: 5.2
IJAR 2015; 1(9): 201-203
www.allresearchjournal.com
Received: 06-06-2015
Accepted: 09-07-2015

E Seetha
M. Phil Research Scholar, Department of Computer Science and Applications. Vivekanandha College of Arts and Sciences for women, Namakkal, Tamil Nadu, India.

Mrs. D Ponniselvi
M.Sc., M. Phil, Assistant Professor, Department of computer Science and Applications, Vivekananda College of Arts and Sciences for women, Namakkal, Tamil Nadu, India.

Correspondence
E Seetha
M Phil Research Scholar, Department of Computer Science and Applications. Vivekanandha College of Arts and Sciences for women, Namakkal, Tamil Nadu, India.

A Survey on Authorized Deduplication in cloud storage environment

E Seetha, D Ponniselvi

Abstract
Cloud computing is the most important emerging computing in which resources are shared over the internet. However, cloud storage environment faces one serious problem such as management of large volume of data. To manage vast amount of data, a deduplication technique is used. Data deduplication is one of the important data compression techniques for eliminating duplicate copies of repeating data. To keep the confidentiality of sensitive data while supporting the deduplication, to encrypt the data before outsourcing convergent encryption technique has been proposed. To better protect data security, this project makes the first attempt to formally address the problem of authorized data deduplication.Different from traditional deduplication systems, the differential privileges of users are further considered in duplicate check besides the data itself. In this paper data de-duplication process will be discussed in detail. In data de-duplication there are several methods available that makes it easy to implement. In this paper we will examine about all methods, processes that are used in data de-duplication. The proposed security models contain the demonstration of security analysis scheme. As a proof of concept, contains the implementation framework of proposed authorized duplicate check scheme and conduct test bed experiments using these prototype. In proposed system contain authorized duplicate check scheme incurs minimal overhead compared to normal operations.

Keywords: Cloud computing, data de-deduplication, repeated data, authorized duplicate check, convergent encryption, confidentiality.

1. Introduction
Cloud computing is a model for delivering information technology services in which resources are retrieved from the internet through web-based interface and application, instead of direct connection to a server. Cloud storage provides a service for the evergreen management of vast amount of data in order to reduce the space and bandwidth. To make reliable and scalable management of data in the cloud computing, deduplication plays a vital role as a conventional technique. De-duplication is a data compression technique which is most commonly used for eliminating repeated copies of data/files in cloud storage to reduce space and bandwidth. This technique is used for reliable storage utilization and to provide scalable network data transfers to reduce number of bytes that must be sent.
Data deduplication may occur as file level as well as block level data de-duplication. Keeping multiple duplicate copies of file/data with similar content, de-duplication detects and eliminates the redundant data by keeping original physical copy. The system recovers the storage consumption and it can be applicable to network data transfer to reduce the number of bytes that must be sent. Data de-duplication stretches lot of reimbursements, refuge and confidentiality anxieties ascend as the users' subtle data is liable to both inside and outside spasms. Profitable cloud storage services such as Dropbox, Mozy and Memopal, have been applying de-duplication for user data to bar preservation cost. Data outsourcing raises security and privacy concerns [3]. Deduplication improves storage and bandwidth competence and is attuned with convergent key management. Traditional encryption requires dissimilar users to encrypt their data with their own key. To stop unauthorized access, a secure proof of possession protocol is additionally required to provide the proof that the user indeed owns the similar file when a duplicate is found. Once the proof of consequent users with the similar file are going to be provided a pointer from the server while not having to transfer the similar file.

1.1. Problem Definition Storage efficiency functions such as de-duplication afford storage providers better utilization of their storage back ends and the ability to serve more customers with the same infrastructure. It is the process by which a storage provider only stores a single copy of a file owned by several of its users and there are four different de-duplication strategies, depending on whether de-duplication happens at the client side (i.e. before the upload) or at the server side, and whether de-duplication happens at a file level or at a block level. De-duplication is most rewarding when it is triggered at the client side, as it also saves upload bandwidth but For these reasons, de-duplication is a critical enabler for a number of popular and successful storage services which offers a cheap, remote storage to the broad public by performing client-side de-duplication, thus it will saving both the network bandwidth and storage costs. Indeed, data de-duplication is arguably one of the main reasons why the prices for cloud storage and cloud backup services have dropped so sharply. As the world moves to digital storage for archival purposes, there is an increasing demand for systems that can provide a secure data storage in a cost-effective manner. By identifying the common chunks of data both within and between files and storing them only once, by this de-duplication can yield cost savings by increasing the utility of a given amount of storage but Unfortunately, de-duplication exploits identical content, while encryption attempts to make all content appear random, when the same content encrypted with two different keys results in very different cipher text. Thus, in encryption combining the space efficiency of de-duplication with the secrecy aspects is problematic. Although data de-duplication brings a lot of benefits to cloud user, security and privacy concerns arise as user sensitive data are susceptible to both inside and outside attacks. While Traditional encryption, providing data confidentiality, is incompatible with data de-duplication. Specifically traditional encryption requires different users to encrypt their data with their own keys. Thus, identical data copies of different users will lead to different cipher texts, which make de-duplication impossible. Thus Convergent encryption has been proposed to enforce data confidentiality while making de-duplication feasible.

2. Security Issues in Cloud

2.1. Security Analysis

The security will be analyzed in terms of two aspects, that is, the confidentiality of data and the authorization of duplicate check. We suppose that all the files are sensitive and needed to be fully protected against both public cloud and private cloud. Under this assumption, two kinds of adversaries are considered, adversaries those aim to extract secret information as much as possible from both public cloud and private cloud, and internal adversaries who aim to obtain more information on the file from the public cloud and duplicate-check token information from the private cloud outside of their scopes. The data will be encrypted in our de-duplication system before outsourcing to the storage cloud to maintain the confidentiality of data. The data is encrypted with the traditional encryption scheme and the data encrypted with such encryption method which guarantees the security of data.

2.2. Privacy Preservation

System address the problem of privacy preserving de-duplication in cloud computing and propose a new de-duplication system supporting for Differential Authorization and unauthorized Duplicate Check. Each authorized user is able to get his/her individual token of his file to perform duplicate check based on his privileges. Under this assumption, any unauthorised user cannot generate a token for duplicate check out of his privileges or without the aid from the private cloud server. Authorized user is able to use his/her individual private keys to generate query for certain file and the privileges he/she owned with the help of private cloud, while the public cloud performs the duplicate check directly and tells the user if there is any duplicate.

3. Literature Survey

3.1. Deduplication Analysis

The propose takes an investigation on personal data by investigate how the space consumption efficiency of chunk data, data redundancy and decrease the computation overhead and also the functioning of the hash function. The study Application aware de-duplication motivates the design by following some interpretation of de-duplication for cloud backup services in the personal computing environment:

- The most of storage space is filled by a small number of compressed files with low sub-file redundancy after file-level de-duplication.
- The maximum arrangement of hash finger printing and chunking method which can helps to reduce the system overhead on resources.
- Data shared between various types of applications is insignificant due to difference in data content and format.
- To achieve high de-duplication efficiency with various application datasets there will be the best choices of chunking methods and deployment of chunk level redundancy.

In this paper QoS (*P. Anderson and L. Zhang et al. 2010)* [1] proposed a solution here the data which is common between users to increase the speed of backup and reduce the storage requirement namely backup algorithm. Supports client-end per user encryption is necessary for confidential personal data. This provides the potential to significantly decrease backup times and storage requirement. Storing huge amount of data in personal computer or laptops causes poor connectivity also may be theft due to hardware failure. However Network bandwidth can be a bottle-neck and Backing up directly to a cloud can be very costly are not addressed. Conventional backup's solutions are not well suited to this environment. So client side dedupe necessary for confidential personal data.

In this paper QoS (*J.R. Douceur, A. Adya, W.J. Bolosky, D. Simon et al.*) The Farsite distributed file system provides availability by replicating each file onto multiple desktop computers. In the view of the fact that this replication consumes considerable storage space, it is essential 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. So there is need to present a mechanism to reclaim space from this incidental duplication to make it available for controlled file replication. Our mechanism includes convergent encryption, which enables duplicate files to combine into the space of a single file, even if the files are encrypted with different users [3].

In this paper QoS (*D. Harnik, B. Pinkas and A. Shulman-Peleg*) [6] Cloud storage services commonly use de-duplication, which eliminates redundant data by storing only

a single copy of each file or block. Deduplication reduces the space and bandwidth requirements of data storage services, and is most effective when applied across multiple users, a common practice by cloud storage offerings. In this context they have demonstrated how de-duplication can be used as a side channel which reveals information about the contents of files of other users. In a different scenario, de-duplication can be used as a covert channel by which malicious software can communicate with its control center, regardless of any firewall settings at the attacked machine. Due to the high savings offered by cross-user de-duplication, cloud storage providers are unlikely to stop using this technology. So they propose simple mechanisms that enable cross-user de-duplication while greatly reducing the risk of data leakage.

In this paper Qos (*M. Mulazzani, S. Schrittwieser, M. Leithner, M. Huber*) [7] throughout the past few years, an enormous number of online file storage services have been introduced. At the same time as several of these services provide basic functionality such as uploading and retrieving files by a specific user, more advanced services offer features such as shared folders, real-time association, and minimization of data transfers or unrestricted storage space. Overviews of existing file storage services and examine Dropbox, an advanced file storage solution, in depth. Based on the results they show that Dropbox is used to store copyright-protected files from a popular file sharing network.

In this paper QoS (*M. Bellare, S. Keelveedhi, and T. Ristenpart*) 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 de-duplication, a goal currently targeted by numerous cloud-storage providers. MLE is a primitive of both practical and theoretical concern [2].

In this paper QoS (*S. Halevi, D. Harnik, B. Pinkas, and A. Shulman- Peleg*) Cloud storage systems are becoming increasingly popular. A technology that keeps their cost down is de-duplication, which stores only a single copy of redundant data. Client-side de-duplication attempts to recognize de-duplication opportunities already at the client and save the bandwidth of uploading copies of existing files to the server. Attacks that exploit client-side de-duplication, allowing an attacker to gain access to arbitrary-size files of other users based on very small hash signatures of these files. More specifically, an attacker who knows the hash signature of a file can convince the storage service that it owns that file; hence the server lets the attacker download the entire file [5].

4. Conclusion

Data Deduplication eradicates the redundant data by storing only the single copies of data. It uses the convergent encryption technique to encrypt the data with Authorized duplicate check, so that only authorized user with specified privileges can perform the duplicate check. The concept de-duplications save the bandwidth and reduce the storage space. It also eradicates the duplicates of data in the cloud storage. It does not allow the unauthorized user to steal information. Thus it provides lots of benefits based on the confidentiality, authorized duplicate check also the cloud storage space as well as the healing information is prevented.

5. References

1. Anderson P, Zhang L. Fast and Secure Laptop Backups with Encrypted De-Duplication in Proc. USENIX LISA, 2010, 1-8.
2. Bellare M, Keelveedhi S, Ristenpart T. Message-Locked Encryption and Secure De-duplication in Proc. IACR Cryptology Print Archive, 2012; 631:296-312.
3. Douceur JR, Adya A, Bolosky WJ, Simon D, Theimer M. ‘‘Reclaiming Space fromDuplicate Files in a Serverless Distributed File System,’’ in Proc. ICDCS, 2002, 617-624.
4. Gantz J, ReinselD. The Digital Universe in 2020: Big Data, Bigger Digital Shadows, [Online] Available: http://www.emc.com/collateral/analystreports/idc-the-digital-universe-in-2020,Dec, 2012.
5. Halevi S, Harnik D, Pinkas B, Shulman- Pelegc A, ‘‘Proofs of Ownership in Remote Storage Systems,’’ in Proc. ACM Conf. Comput. Commun. Security, Y. Chen, G. Danezis, and V. Shmatikov, Eds 2011, 491-500.
6. Harnik D, Pinkas B, Shulman-Peleg A, ‘‘Side Channels in Cloud Services: De-duplication in Cloud Storage,’’ IEEE Security Privacy, vol. 8, no. 6, pp. 40-47, Nov-Dec, 2010.
7. Mulazzani M, Schrittwieser S, Leithner M, Horizon M. Using Cloud Storage as Attack Vector and Online Slack Space,’’ in Proc. USENIX Security, 2011.