

Chapter 40

Secure Data Deduplication of Encrypted Data in Cloud

Sumit Kumar Mahana

National Institute of Technology Kurukshetra, India

Rajesh Kumar Aggarwal

National Institute of Technology Kurukshetra, India

ABSTRACT

In the present digital scenario, data is of prime significance for individuals and moreover for organizations. With the passage of time, data content being produced increases exponentially, which poses a serious concern as the huge amount of redundant data contents stored on the cloud employs a severe load on the cloud storage systems itself which cannot be accepted. Therefore, a storage optimization strategy is a fundamental prerequisite to cloud storage systems. Data deduplication is a storage optimization strategy that is used for deleting identical copies of redundant data, optimizing bandwidth, improves utilization of storage space, and hence, minimizes storage cost. To guarantee the security parameter, the data which is stored on the cloud must be in an encrypted form to ensure the security of the stored data. Consequently, executing deduplication safely over the encrypted information in the cloud seems to be a challenging job. This chapter discusses various existing data deduplication techniques with a notion of securing the data on the cloud that addresses this challenge.

INTRODUCTION

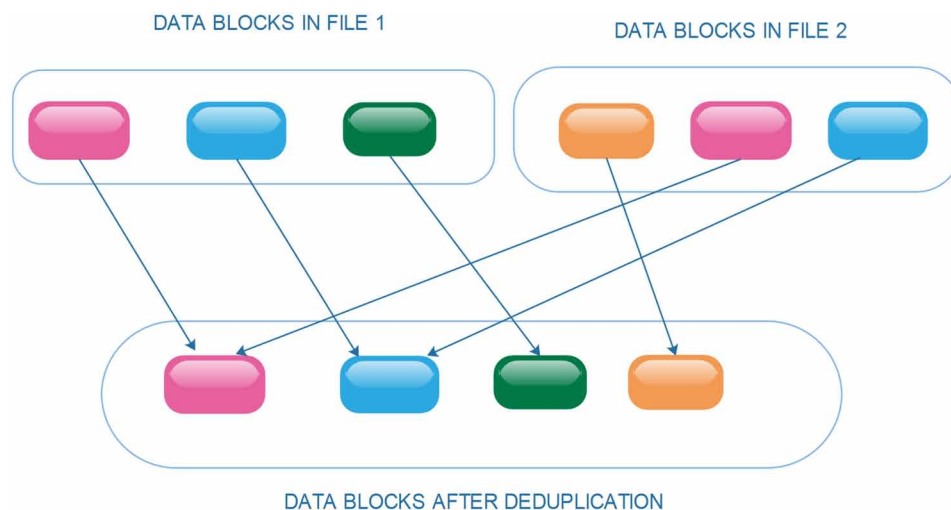
Recent technology improvements have given rise to acceptance and progress of cloud computing. Through various advantages of distributed storage, for instance, ease of access, optimizing bandwidth etc. end users across the globe be likely to move their valuable information to secure distributed platform like cloud storage. According to the report by International Data Corporation (IDC), the huge data content in the whole world will reach around 40 trillion gigabytes by the year 2020 (Chen, Mu, Yang & Guo, 2015; Gantz & Reinsel, 2012). As the quantity of the data content which is being produced continuously increases with the passage of time, there is an urgency to eliminate the duplicate data content which is

DOI: 10.4018/978-1-7998-7705-9.ch040

being stored on the distributed data storage medium like cloud storage to give sufficient storage to the users (Movaliya & Shah, 2016). But, when outsourcing the data to a third party like cloud storage server; it causes the security and privacy issues to turn out to be a serious concern. Distributed storage providers utilize distinctive strategies to enhance storage effectiveness and one of the prominent strategies utilized by numerous cloud storage providers is deduplication, which claims to be saving a huge amount of storage (Douceur, Adya, Bolosky, Simon & Theimer, 2002). Figure 1 illustrates the data deduplication process. Data deduplication, at times also named as intelligent compression or single-instance storage which is frequently used in combination with other forms of reducing data.

Presently, data deduplication technique is broadly utilized by different commercially known cloud storage providers i.e. Dropbox, Amazon S3, Mozy, Google Drive, Memopal etc. to save a lot of storage space and maintenance cost (Akhila, Ganesh & Sunitha, 2016; Di Pietro & Sorniotti, 2012; Harnik, Pinkas, & Shulman-Peleg, 2010).

Figure 1. Data deduplication



WORKFLOW OF DATA DEDUPLICATION

Data deduplication is accomplished in three major steps: chunking, fingerprinting and indexing of fingerprints. The workflow of data deduplication is shown in figure 2. Chunking partitions the input data file stream into small blocks known as chunks. There are mainly two distinct approaches to divide these input data file stream into chunks: fixed size chunking and variable size chunking. The deduplication detection ratio of variable-size chunks is better than fixed-size chunks.

After the process of chunking, some cryptographically secure hash signature (SHA-1, SHA-256, MD5) is applied on chunks to calculate their fingerprints. The process is referred as fingerprinting. There must be a unique Fingerprint value for each chunk. Only unique chunks are placed on the disk after verifying their uniqueness using their fingerprints and redundant chunks are only referenced with old chunks. Indexing is a way to organize fingerprints on disk. Numerous strategies have been suggested for on-disk

15 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/secure-data-deduplication-of-encrypted-data-in-cloud/270630

Related Content

A New Methodology to Arrive at Membership Weights for Fuzzy SVM

Maruthamuthu A., Punniyamoorthy Murugesanand Muthulakshmi A. N. (2022). *International Journal of Fuzzy System Applications* (pp. 1-15).

www.irma-international.org/article/new-methodology-arrive-membership-weights/285556

Game Theory: A Potential Tool for the Design and Analysis of Patient-Robot Interaction Strategies

Aodhan L. Coffey, Tomas E. Wardand Richard H. Middleton (2011). *International Journal of Ambient Computing and Intelligence* (pp. 43-51).

www.irma-international.org/article/game-theory-potential-tool-design/58340

Extracting Functional Dependencies in Large Datasets Using MapReduce Model

K. Amshakala, R. Nedunchezianand M. Rajalakshmi (2014). *International Journal of Intelligent Information Technologies* (pp. 19-35).

www.irma-international.org/article/extracting-functional-dependencies-in-large-datasets-using-mapreduce-model/116741

Statistical Modeling in Healthcare: Shaping the Future of Medical Research and Healthcare Delivery

Mina Bahadori, Morteza Soltani, Masoumeh Soleimaniand Mahsa Bahadori (2023). *AI and IoT-Based Technologies for Precision Medicine* (pp. 431-446).

www.irma-international.org/chapter/statistical-modeling-in-healthcare/332848

An Effective Skin Disease Segmentation Model Based on Deep Convolutional Neural Network

Ginni Arora, Ashwani Kumar Dubeyand Zainul Abdin Jaffery (2022). *International Journal of Intelligent Information Technologies* (pp. 1-18).

www.irma-international.org/article/an-effective-skin-disease-segmentation-model-based-on-deep-convolutional-neural-network/298695