Achieving Efficient and Privacy-Preserving Cross-Domain Big Data Deduplication in Cloud

ABSTRACT

Secure data deduplication can significantly reduce the communication and storage overheads in cloud storage services, and has potential applications in our big data-driven society.

EXISTING SYSTEM

Existing data deduplication schemes are generally designed to either resist brute-force attacks or ensure the efficiency and data availability. The technique of data deduplication is designed to identify and eliminate duplicate data, by storing only a single copy of redundant data. In other words, data deduplication technique can significantly reduce storage and bandwidth requirements .However, since users and data owners may not fully trust cloud storage providers, data (particularly sensitive data) are likely to be encrypted prior to outsourcing. This complicates data deduplication efforts, as identical data encrypted by different users (or even the same user using different keys) will result in different ciphertexts . Thus, how to efficiently perform data deduplication on encrypted data is a topic of ongoing research interest. In recent times, a number of data deduplication schemes have been proposed in the literature. These schemes are designed to realize encrypted data deduplication.For example the _R-MLE2 (Dynamic) scheme proposed in seeks to improve the efficiency of duplicate ciphertext identification. However, the scheme suffers from brute-force attacks, the most popular attack in secure data deduplication schemes Zhou et proposed another efficient secure deduplication scheme Sec Dep to resist brute-force attacks. However, this scheme only deals with small-sized data, and is not suitable for big data deduplication.

DRAWBACKS

- It achieves accountability, in the sense of reducing duplicate information disclosure
- We cannot ensure both efficiency and availability at once.
- It only deals with small-sized data, and is not suitable for big data deduplication.

PROPOSED SYSTEM

In this paper, we investigate a three-tier cross-domain architecture, and propose an efficient and privacy-preserving big data deduplication in cloud storage (hereafter referred to as EPCDD). EPCDD achieves both privacy-preserving and data availability, and resists brute-force attacks. In addition, we take accountability into consideration to offer better privacy assurances than existing schemes. We then demonstrate that EPCDD outperforms existing competing schemes, in terms of computation, communication and storage overheads.

ADVANTAGES

- It eliminates duplicate data.
- It also reduce storage and bandwidth requirements.
- We can ensure both efficiency and availability at once.

SYSTEM REQUIREMENTS

H/W System Configuration:-

➢ Processor	-	Pentium –IV
➢ RAM	-	4 GB (min)
➢ Hard Disk	-	20 GB
➢ Key Board	-	Standard Windows Keyboard
➢ Mouse	-	Two or Three Button Mouse
➢ Monitor	-	SVGA

S/W System Configuration:-

➢ Operating System	:	Linux
➢ Application Server	:	Tomcat5.0/6.X
➢ Backend coding	:	Java
➢ Tool	:	Virtual Box
➢ Environment	:	Ubuntu
➢ Technology	:	Hadoop