Privacy-Preserving Multi-keyword Top-k Similarity Search Over Encrypted Data

ABSTRACT

Cloud computing has emerged as a disruptive trend in both IT industries and research communities recently, its salient characteristics like high scalability and pay-asyou- go fashion have enabled cloud consumers to purchase the powerful computing resources as services according to their actual requirements, such that cloud users have no longer need to worry about the wasting on computing resources and the complexity on hardware platform management. Nowadays, more and more companies and individuals from a large number of big data applications have outsource their data and deploy their services into cloud servers for easy data management, efficient data mining and query processing tasks.

EXISTING SYSTEM

Cloud computing provides individuals and enterprises massive computing power and scalable storage capacities to support a variety of big data applications in domains like health care and scientific research, therefore more and more data owners are involved to outsource their data on cloud servers for great convenience in data management and mining. However, data sets like health records in electronic documents usually contain sensitive information, which brings about privacy concerns if the documents are released or shared to partially untrusted third-parties in cloud. A practical and widely used technique for data privacy preservation is to encrypt data before outsourcing to the cloud servers, which however reduces the data utility and makes many traditional data analytic operators like keyword-based top-*k* document retrieval obsolete.

Disadvantages

- Data privacy preservation is to encrypt data after outsourcing to the cloud servers. So document decryption to authorized data users are not in secure channels.
- Complexity is more on hardware platform management.

PROPOSED SYSTEM

In this paper, we investigate the multi-keyword top-*k* search problem for big data encryption against privacy breaches, and attempt to identify an efficient and secure solution to this problem. Specifically, for the privacy concern of query data, we construct a special tree-based index structure and design a random traversal algorithm, which makes even the same query to produce different visiting paths on the index, and can also maintain the accuracy of queries unchanged under stronger privacy. For improving the query efficiency, we propose a group multi-keyword top-*k* search scheme based on the idea of partition, where a group of tree-based indexes are constructed for all documents. Finally, we combine these methods together into an efficient and secure approach to address our proposed top-*k* similarity search. Extensive experimental results on real-life data sets demonstrate that our proposed approach can significantly improve the capability of defending the privacy breaches, the scalability and the time efficiency of query processing over the state-of-the-art methods.

Advantages

- It shares the secret key of trapdoor generation and document decryption to authorized data users with secure channels.
- The cloud service providers can easily access and analyze the encrypted data and even record queries to learn additional information.

SOFTWARE REQUIREMENTS

Front-end	:	JSP
Back-End	:	MySQL
Server	:	Tomcat Server
OS	:	WINDOWS 7/above

HARDWARE REQUIREMENTS

PROCESSOR	:	CORE i3
RAM	:	512MB-2GB
HARD DISK	:	40GB