**An Efficient Privacy-Preserving Ranked Keyword Search Method**

**ABSTRACT:**

Cloud data owners prefer to outsource documents in an encrypted form for the purpose of privacy preserving. Therefore it is essential to develop efficient and reliable ciphertext search techniques. One challenge is that the relationship between documents will be normally concealed in the process of encryption, which will lead to significant search accuracy performance degradation. Also the volume of data in data centers has experienced a dramatic growth. This will make it even more challenging to design ciphertext search schemes that can provide efficient and reliable online information retrieval on large volume of encrypted data. In this paper, a hierarchical clustering method is proposed to support more search semantics and also to meet the demand for fast ciphertext search within a big data environment. The proposed hierarchical approach clusters the documents based on the minimum relevance threshold, and then partitions the resulting clusters into sub-clusters until the constraint on the maximum size of cluster is reached. In the search phase, this approach can reach a linear computational complexity against an exponential size increase of document collection. In order to verify the authenticity of search results, a structure called minimum hash sub-tree is designed in this paper. Experiments have been conducted using the collection set built from the IEEE Xplore. The results show that with a sharp increase of documents in the dataset the search time of the proposed method increases linearly whereas the search time of the traditional method increases exponentially. Furthermore, the proposed method has an advantage over the traditional method in the rank privacy and relevance of retrieved documents.

**EXISTING SYSTEM:**

* In the recent years, researchers have proposed many ciphertext search schemes by incorporating the cryptography techniques. In addition, the relationship between documents is concealed in the above methods. The relationship between documents represents the properties of the documents and hence maintaining the relationship is vital to fully express a document. For example, the relationship can be used to express its category. If a document is independent of any other documents except those documents that are related to sports, then it is easy for us to assert this document belongs to the category of the sports.
* Due to the blind encryption, this important property has been concealed in the traditional methods. Therefore, proposing a method which can maintain and utilize this relationship to speed the search phase is desirable.
* Sun et al. use Merkle hash tree and cryptographic signature to create a verifiable MDB-tree. However, their work cannot be directly used in our architecture which is oriented for privacy-preserving multiple keyword search. Thus, a proper mechanism that can be used to verify the search results within big data scenario is essential to both the CSPs and end users.

**DISADVANTAGES OF EXISTING SYSTEM:**

* Existing methods have been proven with provable security, but their methods need massive operations and have high time complexity. Therefore, former methods are not suitable for the big data scenario where data volume is very big and applications require online data processing.
* Song et al. method has a high searching cost due to the scanning of the whole data collection word by word.
* Due to the lack of rank mechanism, users have to take a long time to select what they want when massive documents contain the query keyword. Thus, the order-preserving techniques are utilized to realize the rank mechanism,
* Sun et al. give a new architecture which achieves better search efficiency. However, at the stage of index building process, the relevance between documents is ignored.

**PROPOSED SYSTEM:**

* In this paper, a vector space model is used and every document is represented by a vector, which means every document can be seen as a point in a high dimensional space. Due to the relationship between different documents, all the documents can be divided into several categories.
* Instead of using the traditional sequence search method, a backtracking algorithm is produced to search the target documents. Cloud server will first search the categories and get the minimum desired sub-category. Then the cloud server will select the desired k documents from the minimum desired sub-category. The value of k is previously decided by the user and sent to the cloud server. If current sub-category can not satisfy the k documents, cloud server will trace back to its parent and select the desired documents from its brother categories. This process will be executed recursively until the desired k documents are satisfied or the root is reached.
* To verify the integrity of the search result, a verifiable structure based on hash function is constructed.

**ADVANTAGES OF PROPOSED SYSTEM:**

* The search time can be largely reduced by selecting the desired category and abandoning the irrelevant categories.
* Comparing with all the documents in the dataset, the number of documents which user aims at is very small. Due to the small number of the desired documents, a specific category can be further divided into several sub-categories.
* A virtual root is constructed to represent all the data and categories. The virtual root is denoted by the hash result of the concatenation of all the categories located in the first level. The virtual root will be signed so that it is verifiable. To verify the search result, user only needs to verify the virtual root, instead of verifying every document.

**SYSTEM REQUIREMENTS:**

**HARDWARE REQUIREMENTS:**

* System : Pentium Dual Core.
* Hard Disk : 120 GB.
* Monitor : 15’’ LED
* Input Devices : Keyboard, Mouse
* Ram : 1 GB

**SOFTWARE REQUIREMENTS:**

* Operating system : Windows 7.
* Coding Language : JAVA/J2EE
* Tool : Netbeans 7.2.1
* Database : MYSQL

**REFERENCE:**

Chi Chen, Member, IEEE, Xiaojie Zhu, Student Member, IEEE, Peisong Shen, Student Member, IEEE, Jiankun Hu, Member, IEEE, Song Guo, Senior Member, IEEE, Zahir Tari, Senior Member, IEEE, and Albert Y. Zomaya, Fellow, IEEE, “An Efficient Privacy-Preserving Ranked Keyword Search Method”, **IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 27, NO. 4, APRIL 2016.**