Enabling Personalized Search over Encrypted Outsourced Data with Efficiency Improvement

B. Pavan Kumar[1], Dr.E.Kesavulu Reddy[2]
[1]*III YEAR MCA, Dept of MCA SVU College of CM&CS*
[2] *Assistant Professor, Dept of MCA SVU College of CM&CS*

Abstract- **In cloud computing, searchable encryption scheme over outsourced data is a hot research field. However, most existing works on encrypted search over outsourced cloud data follow the model of "one size fits all" and ignore personalized search intention. Moreover, most of them support only exact keyword search, which greatly affects data usability and user experience. So how to design a searchable encryption scheme that supports personalized search and improves user search experience remains a very challenging task. In this paper, for the first time, we study and solve the problem of personalized multi-keyword ranked search over encrypted data while preserving privacy in cloud computing. We build a user interest model for individual user by analyzing the user's search history, and adopt a scoring mechanism to express user interest smartly. To address the limitations of the model of "one size fit all" and keyword exact search, we propose two personalized search schemes for different search intentions. Extensive experiments on real-world dataset validate our analysis and show that our proposed solution is very efficient and effective.**

I. INTRODUCTION

Cloud computing has achieved great development both in academic and industry communities as it provides economic and convenient service. And now more and more companies and users are planning to upload their data onto the public clouds. However, data stored in the cloud may suffer from malicious use by cloud service providers since data owners have no longer direct control over data. Considering data privacy and security, it is a recommended practice for data owners to encrypt data before uploading onto the cloud. Although it protects data security from illegal use both from cloud service providers and external users, it makes data utilization more difficult since many techniques based on plaintext are no longer applicable to cipher text. Therefore, exploring an efficient search technique for encrypted data is extremely urgent. A popular way to search over encrypted data is searchable encryption and many constructive schemes have been put forward under different applications. However, these searchable encryption schemes based on keyword no longer fully satisfy the new challenge and users' increasing needs, specifically manifested in the following two aspects.

In those schemes, the cloud will return all files that match the user's query, which may cause a huge consumption of network bandwidth. Moreover, it will cost user much time and many resources to filter his real interesting ones among a large quantity of returned files. In the practical application, different users may find different things relevant because of different importance or priorities of query terms, indicating the necessity of personalized search, which takes personal keyword preference or keyword priority into account.

II. EXISTING SYSTEM

In order to address the limitation of the model of "one size fits all" in most existing searchable encryption schemes. Due to the fact that most of existing searchable encryption schemes support only exact keyword search, it is possible that the user just gets a few results by querying some terms. Most of existing searchable encryption schemes support only exact keyword search, which affects data usability and user's experience. However, they have not been widely adopted yet since users worry about the leakage of user privacy. Moreover, most of existing personalized search schemes are inapplicable to cipher text.

III. PROPOSED SYSTEM

A preferred keyword search scheme over encrypted data, but the artificial manner of measuring keyword preference has great randomness and fails to consider different users' search histories. Keyword Weight Keywords are practical tools to summarize document content.

Keyword Priority Through the well-trained user interest model, we can get the access frequency of each keyword. The higher the access frequency of a keyword is, the more important the keyword is from viewpoint of the user. The importance of the keyword usually means its rank priority. Relevance Score It is used to measure the score of the query to a document. We can divide the whole relevance score into many sub-scores to represent the connection of file to the keywords in the query.

Secure Inner Product When calculating the relevance score of the document to the query, we should use two vectors: the document vector and the query vector. However, it is not advisable to directly outsource two vectors onto the cloud at the risk of leaking index privacy and query privacy.

IV . IMPLEMENTATION

Implementation is the stage of the project when the theoretical design is turned out into a working system. Thus it can be considered to be the most critical stage in achieving a successful new system and in giving the user, confidence that the new system will work and be effective.

The implementation stage involves careful planning, investigation of the existing system and it's constraints on implementation, designing of methods to achieve changeover and evaluation of changeover methods.

V. USER INTEREST MODEL CONSTRUCTION

In order to evaluate the construction of user interest model, we randomly select three users, analyze their historic records and extract their query terms. The time cost among three users with respect to different size of historic records. We can see the time cost rises with number of historic records with respect to a user. Due to diversity of historic records among the users, their time cost in the same size of historic records varies widely. Besides, we record the storage overhead of user interest model with different size of query terms.

The time cost to generate a query mainly depends on the number of keywords in the dictionary, since the common main operation or timeconsuming operation in all the schemes is query encryption. So the time cost will become large as increasing the number of keywords in the dictionary.

VI. SEARCHABLE ENCRYPTION

The first practical searchable encryption scheme in symmetric setting is proposed by Song et al. in which each word is encrypted independently under a two layer construction and the user has to go through the whole document to search a certain keyword. And then some security definitions and many improvements or constructions have been proposed the first public key-based searchable encryption scheme, where anyone owning the private key can search data items encrypted by the public key. Recently, the problem of information leakage in conjunctive search and dynamic searchable symmetric encryption, respectively. After that, a lot of schemes under different applications have been proposed. Among them, to address the spelling mistake, import issue of fuzzy keyword search is ranking on single keyword and multi-keyword respectively. After that, Sun et al. improve the efficiency of multi-keyword search by adopting MDB-tree. However, most of existing searchable encryption schemes support only exact keyword search, which affects data usability and user's experience. Fu et al. propose a semantic keyword search scheme based on stemming algorithm, which helps users find relevant documents containing semantically close keywords related to the query word. Furthermore, personalized search is also missed or ignored. Proposed a preferred keyword search scheme over encrypted data, but how to measure keyword preference is ignored. The artificial method of measuring keyword preference is time consuming and imposes a burden on the user. Moreover, it fails to consider different users' search histories and thus has great randomness.

VII. PERSONALIZED SEARCH

Personalized search aims at exploiting user information to enable search results better meet the individual user's search intention. The general approach is to build a user profile, which describes the user's interests or preferences that can directly set by the user or collected during the search history for a period. And the most challenging works in personalized search are:

1) How to build a user profile model

2) How to make use of the user profile to improve the search.

The user profile model is often built upon a set of keyword vectors or classes of keyword vectors. Due to the absence of interrelation of keywords in the keyword-based representation model, some researchers make use of a set of concepts derived from predefined ontology or reference ontology to express the user profile model. It is well known that ontology-based user profile model is superior to other representation models as it can exploit semantic knowledge. Note that search personalization is achieved by integrating the user profile in the query reformulation process, query-document matching , personalized document categorization or document re-ranking. The current search engines, such as GOOGLE, has launched personalized search, where the user can indicate his interests or preferences explicitly, or preferences may be automatically gathered. However, they have not been widely adopted yet since users worry about the leakage of user privacy. Moreover, most of existing personalized search schemes are inapplicable to cipher text.

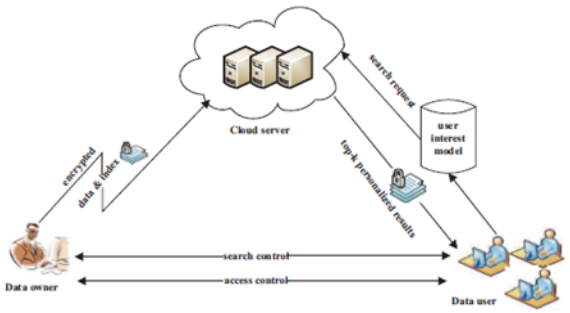

Fig. 1: Architecture of the search over encrypted cloud data

IX. CONCLUSION

we address the problem of personalized multi-keyword ranked search over encrypted cloud data. Considering the user search history, we build a user interest model for individual user with the help of semantic ontology Word Net. Through the model, we have realized automatic evaluation of the keyword priority and solved the limitation of the artificial method of measuring.

Moreover, we propose two PRSE schemes to solve two limitations (the model of "one size fit all" and keyword exact search) in most existing searchable encryption schemes. In addition, through privacy analysis and performance analysis demonstrates that our scheme is practicable.