Searchable Encryption Using a Dynamic Multi-keyword Ranked Search over Cloud Data

Jayapriya Selvam, PG Scholar
Department of Computer Science and Engineering
K. Ramakrishnan College of Technology
Trichy, Tamilnadu, India
sjpriya25@gmail.com

Matheswaran P, Assistant Professor
Department of Computer Science and Engineering
K. Ramakrishnan College of Technology
Trichy, Tamilnadu, India
mathesh3@gmail.com

***Abstract* - Cloud computing has emerging as a promising pattern for data outsourcing and high-quality data services. However, concerns of sensitive information on cloud potentially cause privacy problems. Data encryption protects data security to some extent, but at the cost of compromised efficiency. To tackle this issue, present a secure multi-keyword ranked search scheme over encrypted cloud data, which simultaneously supports dynamic update operations. Like deletion and insertion of documents. Specifically, the vector space model and the widely-used TF_IDF model are combined in the index construction and query generation. To construct a special tree-based index structure and propose a "Greedy Depth-first Search" algorithm to provide efficient multi-keyword ranked search. The secure kNN algorithm is utilized to encrypt the index and query vectors, and meanwhile ensure accurate relevance score calculation between encrypted index and query vectors. In order to resist statistical attacks, phantom terms are added to the index vector for blinding search results. Due to the use of our special tree-based index structure, the proposed scheme can achieve sub-linear search time and deal with the deletion and insertion of documents flexibly.**

Keywords— Multi Keyword search, Ranking, Indexing, Classification, Access control

I. INTRODUCTION

Cloud computing is the practice of using a network of remote servers hosted on the Internet to store, manage, and process data, rather than a local server or a personal computer. It contains potential privacy risks which lead to privacy and security concerns because data is travelling over the internet and is stored in multiple redundant offsite locations available on the internet, which allows application software to be operated using internet enable devices. It is a flexible, cost-effective and proven delivery platform for providing business or consumer IT services over the Internet. Security is one of the most crucial aspects among those prohibiting the wide-spread adoption of cloud computing because essential services are often outsourced to a third party, which makes it harder to maintain data security and privacy, support data and service availability .

The cloud service providers (CSPs) that keep the data for users may access users' sensitive information without authorization. A general approach to protect the confidentiality is to encrypt the data before outsourcing. However, this will cause a huge cost in terms of data usability. For example, the existing techniques on keyword-based information retrieval, which are widely used on the plaintext data, cannot be directly applied on the encrypted data. Downloading all the data from the cloud and decrypt locally is obviously impractical.

Researchers have designed some general purpose solutions with fully-homo morphic encryption or oblivious RAMs. However, these methods are not practical due to their high computational overhead for both the cloud sever and user. On the contrary, more practical special- purpose solutions, such as searchable encryption (SE) schemes have made specific contributions in terms of efficiency , functionality and security. Searchable encryption schemes enable the client to store the encrypted data to the cloud and execute keyword search over cipher text domain. So far, abundant works have been proposed under different threat models to achieve various search functionality, such as single keyword search, similarity search, multi-keyword Boolean search, ranked search, multi-keyword ranked search, etc.Your goal is to simulate the usual appearance of papers in *IEEE conference proceedings*. For items not addressed in these instructions, please refer to the last issue of your conference's proceedings for reference or ask your conference Publications Chair for instructions.

II. MULTI-KEYWORD SEARCHING

Multi-keyword ranked search achieves more and more attention for its practical applicability. Recently, some dynamic schemes have been proposed to support inserting and deleting operations on document collection. These are significant works as it is highly possible that the data owners need to update their data on the cloud server. But few of the dynamic schemes support efficient multi- keyword ranked search.

To proposes a secure tree-based search scheme over the encrypted cloud data, which supports multi- keyword ranked search and dynamic operation on the document collection. Specifically, the vector space model and the widely used "term frequency (TF)×inverse document frequency (IDF)" model are combined in the index construction and query generation to provide multi- keyword ranked search. In order to obtain high search efficiency, we construct a tree-based index structure and propose a "Greedy Depth-first Search" algorithm based on this index tree. Due to the special structure of our tree-based

index. The balanced binary tree is widely used to deal with optimization problems . Finding the best solution from all feasible solution. The keyword balanced binary (KBB) tree in our scheme is a dynamic data structure. The proposed search scheme can flexibly achieve sub-linear search time and deal with the deletion and insertion of documents. The secure kNN algorithm is utilized to encrypt the index and query vectors, and meanwhile ensure accurate relevance score calculation between encrypted index and query vectors. Ranked search can enable quick search of the most relevant data. Sending back only the top-k most relevant documents can effectively decrease network traffic.

Unencrypted dynamic multi-keyword ranked search (UDMRS) scheme which is constructed on the basis of vector space model and KBB tree. Based on the UDMRS scheme, we construct two secure search schemes: the basic dynamic multi-keyword ranked search (BDMRS) scheme in the known ciphertext model, and the enhanced dynamic multi-keyword ranked search (EDMRS) scheme in the known background model. Privacy requirements are summarized through as Index confidentiality and query confidentiality, Trapdoor unlinkability, Keyword privacy.

III. MULTI-KEYWORD SEARCHING WORKING

Figure 1.1 describes the system architecture. The system model in this paper involves three different entities: data owner, data user and cloud server.

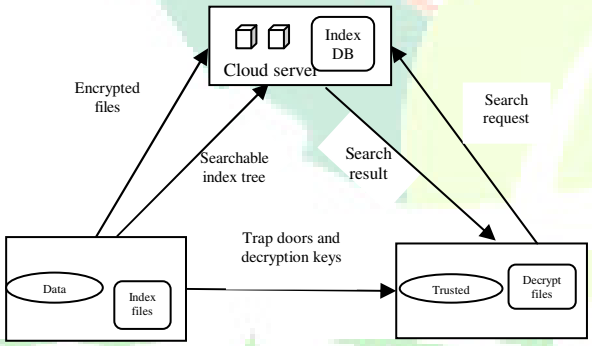

Figure 1.1 System Architecture

Data owner has a collection of documents $F = \{f1; f2; :::; fn\}$ that he wants to outsource to the cloud server in encrypted form while still keeping the capability to search on them for effective utilization. In our scheme, the data owner firstly builds a secure searchable tree index I from document collection F, and then generates an encrypted document collection C for F. Afterwards, the data owner outsources the encrypted collection C and the secure index I to the cloud server, and securely distributes the key information of trapdoor generation (including keyword IDF values) and document decryption to the authorized data users. Besides, the data owner is responsible for the update operation of his documents stored in the cloud server. While updating, the data owner generates the update information locally and sends it to the server.

Data users are authorized ones to access the documents of data owner. With t query keywords, the authorized user can generate a trapdoor TD according to search control mechanisms to fetch k encrypted documents from cloud server. Then, the data user can decrypt the documents with the shared secret key.Cloud server stores the encrypted document collection C and the encrypted searchable tree index I for data owner. Upon receiving the trapdoor TD from the data user, the cloud server executes search over the index tree I, and finally returns the corresponding collection of top- k ranked encrypted documents. Besides, upon receiving the update information from the data owner, the server needs to update the index I and document collection C according to the received information.

IV. SYSTEM DESCRIPTION

The various components used here are: (A) Cloud Storage Construction Module, (B) Data Secure Module, (C) Trapdoor/Index Tree Generation Module, (D) Secure Search Module, (E) Retrieve/ Decrypt Module. The working of each of the components is elaborated in the following section:

A. Cloud Storage Construction Module

The data owner takes a security parameter and outputs invertible matrixes as well as a dimension binary vector S as the secret key, where d represents the size of the keyword dictionary. Then, the data owner generates a set of attribute keys sk for each search user according to her role in the system. The data owner chooses a key KT for a symmetric cryptography Enc ().

The data owner chooses a full-domain collusion resistant hash function, a full-domain pseudorandom function, a pseudorandom generator and a hash function on the AES block-cipher. Then, the data owner chooses a number $\alpha > 1$ that defines the expansion parameter and a number k that denotes the minimum number of blocks in a communication.

B. Data Secure Module

The data owner builds the secure data encrypted as follows: The data owner computes the d-dimension relevance vector p, for each document using the TF-IDF weighting technique. The data owner extends the p to a (dC2)-dimension vector. For each document di, to compute the encrypted relevance vector, the data owner encrypts the associated extended relevance vector p using the secret key $M1$, $M2$ and S. The each document is combine different blocks. The each block is a header, the header indicating that the blocks belongs to which document. Data owner select Enc() function and encrypt the document.

C. Trapdoor/Index Tree Generation Module

The search user takes a Multi keyword from data owner and generates vector score, select two different keyword from received multi keyword after that encrypt the trapdoor and request to cloud for encrypted formatted. The search user sends *Q*, *stag* and a number *k* to the cloud server to request the most *k* relevant documents. The KBB index tree structure is an index tree, which assists us in introducing the index construction. In the process of index construction, first generate a tree node for each document in the collection. These nodes are the leaf nodes of the index tree. Then, the internal tree nodes are generated based on these leaf nodes.

D. Secure Search Module

Once receiving *Q*, *stag*, and *k*, the cloud server parses the *stag* to get a set of integers in the range of document. Then, the cloud server accesses index z in the blind storage and retrieve the blocks indexed. These blocks consist of the blocks and some dummy blocks. For each retrieved encrypted relevance vector *P*, compute the relevance scorefor the associated document di with the encrypted query vector Q. After sorting the relevance scores and send back of the top-k document that is most relevant to the searched keywords.

The search process of the UDMRS scheme is a recursive procedure upon the tree, named as "Greedy Depthfirst Search (GDFS)" algorithm. Construct a result list denoted as *RList*, whose element is defined as *(RScore; FID)*. Here, the *RScore*is the relevance score of the document *fFID*to the query, which is calculated according to formula . The *RList*stores the *k* accessed documents with the largest relevance scores to the query. The elements of the list are ranked in descending order according to the *RScore*, and will be update timelyduring the search process.

E. Retrieve / Decrypt Module

The search user's attributes satisfy the access policy of the document, the search user can decrypt the descriptor using his secret attribute keys to get the document id and the associated symmetric key. The header id using for recovering the first blocks of the relevance data order and identify the blocks size.

V. CONCLUSION

A secure, efficient and dynamic search scheme is proposed, which supports not only the accurate multi-keyword ranked search but also the dynamic deletion and insertion of documents. We construct a special keyword balanced binary tree as the index, and propose a "Greedy Depth-first Search" algorithm to obtain better efficiency than linear search. In addition, the parallel search process can be carried out to further reduce the time cost. The security of the scheme is protected against two threat models by using the secure kNN algorithm.

There are still many challenge problems in symmetric SE schemes. In the proposed scheme, the data owner is responsible for generating updating information and sending them to the cloud server. It could be a meaningful but difficult future work to design a dynamic searchable encryption scheme whose updating operation can be completed by cloud server only. Actually, there are many secure challenges in a multi-user scheme. Firstly, all the users usually keep the same secure key for trapdoor generation in a symmetric SE scheme. In this case, the revocation of the user is big challenge. If it is needed to revoke a user in this scheme, we need to rebuild the index and distribute the new secure keys to all the authorized users. Secondly, symmetric SE schemes usually assume that all the data users are trustworthy. It is not practical and a dishonest data user will lead to many secure problems. For example, a dishonest data user may search the documents and distribute the decrypted documents to the unauthorized ones. Even more, a dishonest data user may distribute his/her secure keys to the unauthorized ones. In the future works, we will try to improve the SE scheme to handle these challenge problems.

REFERENCES

[1] Kuzu M., Islam M.S., and Kantarcioglu M. (2012), 'Efficient similarity search over encrypted data', *in Data Engineering (ICDE), IEEE 28th International Conference on. IEEE*, pp. 1156–1167.

[2] Li J., Wang Q., Wang C., Cao N., Ren K., and Lou W. (2010), 'Fuzzy keyword search over encrypted data in cloud computing',*in INFOCOM, Proceedings IEEE. IEEE*, pp. 1–5.

[3] Song D.X., Wagner D., and Perrig A. (2000), 'Practical techniques for searches on encrypted data', *in Security and Privacy, S&P Proceedings. IEEE Symposium on. IEEE*, pp. 44–55.

[4] Sun W., Wang B., Cao N., Li M., Lou W., Hou Y.T., and Li H. (2013), 'Privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking', *in Proceedings of the 8th ACM SIGSAC symposium on Information, computer and communications security*. pp. 71–82.

[5] Wang C., Cao N., Ren K., and Lou W.(2012), 'Enabling secure and efficient ranked keyword search over outsourced cloud data', *Parallel and Distributed Systems, IEEE Transactions on, vol. 23, no. 8*, pp. 1467–1479.

[6] Wang B., Yu S., Lou W., and Hou Y.T(2014), 'Privacy-preserving multikeyword fuzzy search over encrypted data in the cloud', *in IEEE INFOCOM*.