Analysis of Central Keyword-based Semantic Extension Search Scheme over Encrypted Outsourced Data

Dipti D. Mehare1, Prof. A. V. Deorankar2

P. G. Scholar, Department of Computer Science,
Government college of Engineering, Amravati, Maharashtra, (India)
Assistant Professor, Department of Computer Science,
Government college of Engineering, Amravati, Maharashtra, (India)

ABSTRACT

Searching keywords in documents have quite different importance when users take search operations and such keywords may have a certain grammatical relationship among them which reflect the importance of keywords from the user's perspective intuitively. The existing search techniques regard the search keywords as independent and unrelated. In this paper for the first time takes the relationship among keywords into consideration. Further, we design a novel central keyword semantic extension ranked scheme. By extending the central query keyword instead of all keywords, my scheme makes a good tradeoff between the search functionality and efficiency. My work first gives a basic idea for the design of the central keyword semantic extension ranked scheme, and then presents two secure searchable encryption schemes to meet different privacy requirements under two different threat models. Experiments on the real-world dataset show that my proposed schemes are efficient, effective and secure.

Keywords: *Keyword search, semantic search, cloud computing*.

I. INTRODUCTION

Cloud computing is a technology, which is used to store data into the cloud on demand high quality applications and services. Some of data stored on cloud is sensitive must be protected from unauthorized access including from cloud operators. A solution for this is for users to encrypt their data before sending them to the storage server. This solution protects the data but is at odds with the utility of the cloud. For effective search over encrypted data, the data owner first builds an encrypted index based on the extracted keywords from data files and the corresponding index-based keyword matching algorithm, and then outsources both the encrypted data and the index structure to the cloud server. To search over the encrypted files, the cloud server integrates the trapdoors of keywords with the index information and finally returns the target files to the data users.

In this paper, we take the relation among query keywords into consideration and design a keyword weighting algorithm to show the importance of distinction of the keywords. Using the keyword weights, we can accurately and efficiently localize the central keyword that the user is interested in. Since we can choose the central keyword of the query to extend, our scheme can greatly reduce the trapdoor generation time. Our work first gives a basic concept for the central keyword semantic extension ranked scheme, and then proposes two secure searchable encryption schemes to meet different privacy requirements under two different threat models.

II. RELATED WORK

A. Multi-keyword Searchable Encryption

Yu et al. proposed a secure multi-keyword top-K retrieval scheme. Their scheme used the VSM and the homomorphic encryption to achieve the top-K retrieval. Xia et al. designed a secure and dynamic multi-keyword search scheme. Their scheme enables the deletion and insertion of documents. However, their scheme is only secure in the known cipher text model. Fu et al. proposed a multi-keyword ranked search scheme supporting parallel search process. Specifically, they used Vector Space Model (VSM) to design a tree-based index structure supporting parallel search, which takes the advantage of powerful computing capacity.

B. Fuzzy Keyword Searchable Encryption

Li et al. utilized edit distance to build a predefined fuzzy keyword set to solve the fuzzy keyword search, but their scheme only supports the single keyword fuzzy search. Then Liu et al. reduced the index size based on Chuah et al. treated the pre-defined phrases as a single keyword and designed a privacy-aware bed tree based multi-keyword fuzzy search scheme.

C. A Secure and Dynamic Multi-Keyword Ranked Searchable Encryption

Z. Xia, X. Wang, X. Sun, and Q. Wang 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. They 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.

D. Searchable Encryption over Feature-Rich Data

Storage services allow data owners to store their huge amount of potentially sensitive data, such as audios, images, and videos, on remote cloud servers in encrypted form. To enable retrieval of encrypted files of interest, many searchable symmetric encryption (SSE) schemes have been proposed by Q. Wang, M. He, M. Du, S. S. M. Chow. However, most existing SSE solutions construct indexes based on keyword-file pairs and focus on boolean expressions of exact keyword matches. Moreover, most dynamic SSE solutions cannot achieve forward privacy and reveal unnecessary information when updating the encrypted databases**.**

III. EXISTING SYSTEM

In existing system, when Data owner outsources his document on cloud, the document will get encrypted using AES and stored on cloud. After encryption when any other user wants to search that document, he will specify search query and trapdoor key. The search query will get processed to extract keywords. The keywords will be searched in every encrypted documents within the specified scope(document group as per the trapdoor key) to calculate document wise keywords weight(frequency).Keywords weight according search result re-ranking will be done. Finally search result will be delivered to user. If user wants to download document, he have to specify secrete key. If secrete key is verified, documents will be decrypted and delivered it to user.

IV. PROPOSED SYSTEM

The proposed system is implemented on well known and widely used cloud computing, we need a lot of space to store our sensitive information and data. We are placing or putting more and more sensitive information to the cloud server, but at the same time security of our data must be strong for ensuring data privacy. Sensitive data have to be encrypted before outsourcing this makes the effective data utilization as a challenge for the user as well as First, none of the existing solutions take the semantic relationship among query keywords into consideration. They regard the input keywords as independent and irrelevant. In fact, the importance of a keyword is quite different from that of the others, and this can be clearly shown by the semantic relations among the query keywords. Another problem is that the existing keyword-based search techniques can only return files that contain the exact query keyword, and are unable to hit the files which contain semantic-relevant keywords. For the first time we formalize and solve problem of semantic or extension keyword search over encrypted cloud data while maintaining keyword privacy. We therefore take the relationship among query keywords into consideration and design a keyword weighting algorithm to show the importance of distinction of the keywords. Using the keyword weights, we can accurately and efficiently localize the central keyword that the user is interested in. Since we can choose the central keyword (not all keywords) of the query to extend, our scheme can greatly reduce the trapdoor generation time. In this way, our scheme makes a good tradeoff between the functionality and the efficiency.

In proposed system, when Data owner outsources his document on cloud, system will read the complete document and apply NLP on sentences to fetch keywords from the document. The keywords fetched from the document will be saved in XML file. Then the document will be encrypted using Advanced AES encryption algorithm and stored on cloud. The Advanced AES algorithm is a user defined algorithm, which divides the document into two parts .The two parts will get reversed and encrypted separately using different keys and AES algorithm. The separately encrypted parts will be merged and stored on cloud. After encryption when any other user wants to search that document, he will specify search query and trapdoor key. The search query will get processed to extract keywords. Synonyms for extracted keywords will be searched over www using any API. Keywords set will improved with available synonyms. The keywords will be searched in XML file rather than document which reduce search time. Keywords weights already saved in XML so there is no need to calculate the weight every time. Keywords weight according search result re-ranking will be done. Finally search result will be delivered to user. If user wants to download document, he have to specify secrete key. If secrete key is verified, documents will be decrypted and delivered it to user. Following diagram shows the system model for central keyword based semantic extension search scheme.

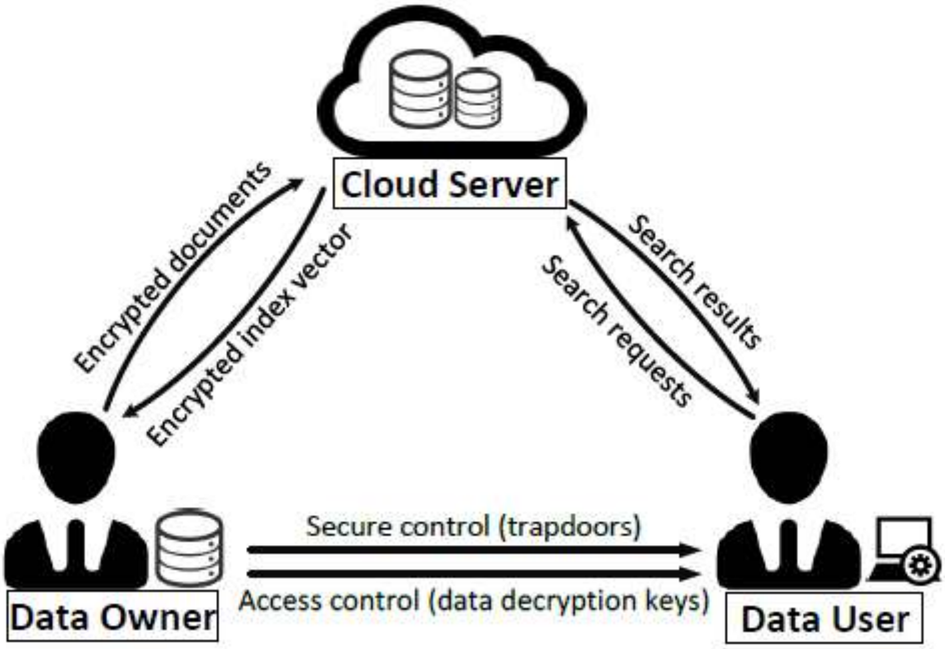

Fig. no. 1: The System Model

V. CONCLUSIONS

In this paper, for the first time, took the relationship among the query keywords into consideration and designed a keyword weighting algorithm based on the relations. Also designed a central keyword semantic extension scheme according to the keyword weights. By choosing the central keyword instead of not all the keywords to extend, our scheme achieves a tradeoff between functionality and efficiency. Moreover, this scheme can support efficient additions of the keyword collection. Finally, proposed two secure searchable encrypted schemes to meet different privacy requirements and evaluated the performance of our schemes comprehensively on a real world dataset.

REFERENCES

[1]. Zhangjie Fu, Xinle Wu, Qian Wang, Kui Ren, Fellow, IEEE," Enabling Central Keyword-based Semantic Extension Search over Encrypted Outsourced Data", IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY.

[2]. Mr. Mahesh Lanjewar , Swapnali Ghadge, Sneha Mane, Priti Dalvi," Fuzzy Keyword Search Over Encrypted Data Using Cloud Computing", International Journal of Engineering Research and Applications (IJERA).

[3]. Tarik Moataz, Abdullatif Shikfay, Nora Cuppens-Boulahia, and Fr´ed´eric Cuppens," Semantic Search Over Encrypted Data".

[4]. Avani Konda, Sai Praneeth Gudimetla, Balaji T, Gopi Krishna Subramanyam, Usha Kiruthika," Synonymous Keyword Search Over Encrypted Data in Cloud", International Journal of MC Square Scientific Research Vol.9, No.2 2017.

[5]. Abinaya. S, Dr. R. Kalpana," Secure and Dynamic Image Search for Retrieving Document" International Journal of Computer Science and Mobile Computing.