R. Geetha *et al*, International Journal of Computer Science and Mobile Applications,
Vol.4 Issue. 4, April- 2016, pg. 01-05 **ISSN: 2321-8363**
Impact Factor: 4.123

EFFICIENT MULTIKEYWORD RANKED SEARCH OVER ENCRYPTED CLOUD DATA WITH RANK INTEGRITY

R. Geetha[*1], M. Padma[#2]

[1]M.E (Mobile and Pervasive Computing), Anna University, BIT-Campus, Trichy

[2]Teaching Fellow, Department of CSE/IT, Anna University, BIT-Campus, Trichy

ABSTRACT: ***Due to the high popularity of cloud computing, more data owners are motivated to outsource the data to the cloud server. In that sensitive data will be encrypted before outsourcing to the cloud server for security purpose. In this paper, we introduce a secure multi-keyword ranked search over encrypted cloud data, which performs dynamic update operations like deletion and insertion of documents. By combining the vector space model and widely used TFxIDF model for the index construction and query generation. By constructing a special tree-based index structure and introduce "Greedy Depth First Search" algorithm that gives effective multi-keyword ranked search. Secure KNN algorithm is used to encrypt the index and query vectors, and also gives accurate relevance score calculation between encrypted index and query vectors. Due to the special tree-based index structure, it can achieve sub-linear search time and perform deletion and insertion of documents flexibly. Using multi-keyword ranked search over encrypted cloud data the files can be retrieved based on ranking. Thus, the ranking provides similar files from the cloud server it cannot assure that retrieved files are same or not. In this paper ranking is tested to the correctness of its order. The Rank test method is used to check the integrity of the rank order of the search results. Since the rank is fixed by the cloud server is tested and the user can get accurate results and so privacy can be improved.***

Keywords: Multi-keyword, rank integrity, index, query, dynamic update, encryption

I. INTRODUCTION

Cloud computing is a new model of enterprise IT infrastructure, which can arrange huge resource of computing, storage, and applications, and enable users to have ubiquitous, convenient and on-demand network access to a shared pool of configurable computing resources with great efficiency and minimal economic overhead . Attracted by these appealing features, both individuals and enterprises are motivated to outsource their data to the cloud, instead of purchasing software and hardware to manage the data themselves. Despite the various advantages of cloud services, outsourcing sensitive information to remote servers brings privacy concerns. The cloud service providers (CSPs) that carry the data for users may access users' sensitive information without authorization. A general approach to keep the data confidentiality is to encrypt the data before outsourcing. However, this will cause a huge cost in terms of data usability. The existing techniques for keyword-based information retrieval are used on the plaintext data, cannot be directly applied to the encrypted data. Downloading all the data from the cloud and decrypting it is obviously impractical. In order to overcome the above problem, researchers have designed some solutions with fully-homomorphic encryption. However, these methods are not practical due to their high computational overhead for both the cloud server 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 a keyword search over the cipher text domain. So far, abundant works have been proposed under various methods to achieve search functionality, such as single keyword search, similarity search, multi-keyword Boolean search, ranked search, multi-keyword ranked search, etc. Among them, multi- keyword ranked search achieves more and more attention for its practical applicability. Dynamic update schemes have been proposed to support inserting and deleting operations on document collection.

But few of the dynamic schemes support efficient multi-keyword ranked search our contributions are summarized as follows:

1) We design a searchable encryption scheme that supports both the accurate multi-keyword ranked search and flexible dynamic operation on document collection.

2) Due to the special structure of tree-based index, the search complexity of the proposed scheme is fundamentally kept to logarithmic.

And in practice, the proposed scheme can do higher search efficiency by executing our "Greedy Depth-first Search" algorithm. Moreover, the parallel search can be flexibly performed to further reduce the time cost of the search process. An effective and flexible Audit scheme support to reduce the computation overheads. To ensure the correctness of similar data rank over the cloud that allowing a third party auditor (TPA), on behalf of the cloud client, to checking the integrity of the rank order in the search result assuming the cloud server is untrusted.

II. RELATED WORK

Searchable encryption schemes enable the clients to store the encrypted data to the cloud and execute keyword search over cipher text domain. Due to different cryptography primitives, searchable encryption schemes can be constructed using public key based cryptography or symmetric key based cryptography Song *et al* proposed the first symmetric searchable encryption (SSE) scheme, and the search time of their scheme is linear to the size of the data collection. Goh proposed formal security definitions for SSE and designed a scheme based on Bloom filter. The search time of Goh's scheme is $O(n)$, where n is the cardinality of the document collection. Curtmola *et al.* proposed two schemes (SSE-1 and SSE-2) which achieve the optimal search time. Their SSE-1 scheme is secure against chosen-keyword attacks (CKA1) and SSE-2 is secure against adaptive chosen-keyword attacks (CKA2). These early works are single keyword Boolean search schemes, which are very simple in terms of functionality. Afterward, 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 and multi-keyword ranked search etc. Multi-keyword Boolean search allows the users to input multiple query keywords to request suitable documents. Among these works, conjunctive keyword search schemes only return the documents that contain all of the query keywords. Disjunctive keyword search schemes return all of the documents that contain a subset of the query keywords. Predicate search schemes are proposed to support both conjunctive and disjunctive search. All these multi keyword search schemes retrieve search results based on the existence of keywords, which cannot provide acceptable result ranking functionality. 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. Some early works have realized the ranked search using order-preserving techniques, but they are designed only for single keyword search. Cao *et al*. realized the first privacy-preserving multi-keyword ranked search scheme, in which documents and queries are represented as vectors of dictionary size. With the "coordinate matching", the documents are ranked according to the number of matched query keywords. However, Cao *et al.*'s scheme does not consider the importance of the different keywords and thus is not accurate enough. In addition, the search efficiency of the scheme is linear with the cardinality of the document collection. Sun *et al.* presented a secure multi-keyword search scheme that supports similarity-based ranking. The authors constructed a searchable index tree based on vector space model and adopted cosine measure together with TF×IDF to provide ranking results. Sun *et al.*'s search algorithm achieves better-than-linear search efficiency but results in the precision loss. O¨ Rencik *et al.* proposed a secure multi-keyword search method which utilized local sensitive hash (LSH) functions to cluster the similar documents. The LSH algorithm is suitable for similar search but cannot provide an exact ranking. The Index Hash Table may increase the complexity of an audit system, it provides a higher assurance to monitor the behavior of an untrusted CSP, as well as valuable evidence for computer forensics, due to the reason that anyone cannot forge the valid Xi (in TPA) and σi (in CSP) without the secret key SK. To support dynamic data auditing operations, to introduce a simple IHT to record the changes of every metadata blocks, as well as generate the hash value of each file in the verification process. The structure of our IHT is similar to that of file block allocation table in file systems. Generally, the IHT consists of a serial number, file number, version number, and rank value. In this module with "whole" checking, rank checking greatly reduces the workload of audit services, while still achieves an effective detection of misbehaviors. Thus, a probabilistic audit on rank checking is preferable to realize the anomaly detection in a timely manner and finally to check the result correct or not through TPA analysis. This auditing process will improve efficient when an auditing the rank and perfectly to measure untrusted cloud returned result over the high-performance way.

III. SYSTEM MODEL

The system model can be divided into three entities, as shown:

Data owner:

Data owner collects the N number of important documents and outsources them to the cloud server. Before outsourcing them the documents are encrypted and uploaded to the cloud server for privacy requirements and securely distribute the key information by trapdoor to the authorized users.

Data user:

They are the authorized ones to access the documents uploaded by the data owner. Using the keyword the authorized user can send a trapdoor according to the search to get the encrypted documents from the cloud server. Then the user can decrypt the documents using a key.

Cloud server:

Cloud server stores the encrypted document *and* the encrypted tree index for data owner. Upon receiving the trapdoor *TD* from the data user, the cloud server executes the search over the index tree, 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 and document collection *C* according to the received information.

IV. EXISTING SYSTEM

Existing searchable encryption schemes provide efficiency, functionality, and security. This scheme enable the client to store the encrypted data to the cloud and retrieve them using keyword search over cipher text domain. Abundant works have been done under different threat models to various search functionality such as single keyword search, similarity search, multi-keyword search, multi-keyword Boolean search, ranked search, multi-keyword search, etc. Among them, multi-keyword ranked search achieves more attention of document retrieval. This paper propose a secure tree based index structure and propose a "Greedy Depth First Search" algorithm to provide efficient multi-keyword ranked search. Secure kNN algorithm is used to encrypt the index and query vector and provide accurate relevance score calculation between encrypted index and query vectors.

V. SYSTEM ARCHITECTURE

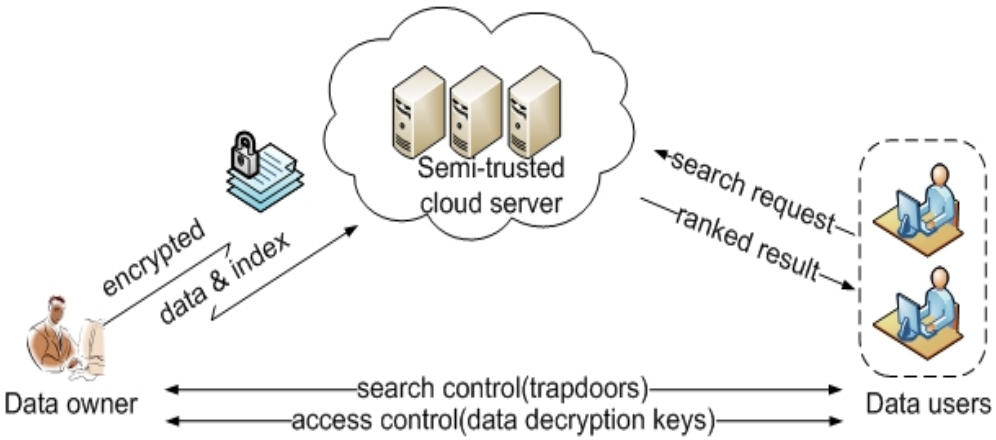

Figure 5.1 Existing system Architecture

VI. PROPOSED SYSTEM

The enormous number of on-demand data users an enormous amount of outsourced data documents are present in the cloud, so this problem is challenging as it is really difficult to meet the requirements of scalability, performance and system usability. Ranked search will only send back the most relevant data hence it can also eliminate unnecessary network traffic, in "pay-as-you-use" cloud paradigm which is highly desirable. There are chances of attacking the rank order sent by cloud server hence it is required to check the integrity of rank. To enhance the accuracy of the search result as well as to improve the user searching experience, it is necessary to support multiple keywords search with the integrated ranking system.

- **Setup** – here by taking the security parameters as an input from data owners generating symmetric keys for security is an output to data owners.

- **Build Index** – here by taking all the documents from the data owners building an index for each and every document by considering unique keywords from the Documents. After index construction, all the keywords and the documents are encrypted before outsourcing into the cloud server.

- **Trapdoor** – here with the keywords of interest as input, this has to generate trapdoor keys for searching the encrypted cloud data.

- **Query** – here data user sends the search query to a cloud server which consists of multiple keywords and trapdoor. It performs keywords to search on the Encrypted cloud for the ranked order of the searchable index.
- **Integrated Rank** – here it checks the integrity of the rank by using hash mapping technique for the integrated ranked ordered search .The sha-1 algorithm is used for checking the integrity of rank. Preserving the integrity of rank is important because when the data user searches for top-k retrieval from the cloud server and when the cloud server returns back the top-k retrieval to the data user there are possibilities of getting attacked by the intruder, so the data user may get the inaccurate rank order. Hence to overcome this problem checking the integrity of rank is required of both cloud and data user. When cloud server sending the top-k retrieval rank order, using rank order it performs a hash mapping technique to generate the signature. Then it sends both rank order and signature to the data user. Now the data user by using the rank order generates signature by using same hash mapping technique. Now the data user verifies the signature which is generated by itself and sent by cloud server if they both matches then the received rank order is accurate else it is inaccurate. If the rank order is accurate user download the document or else discard them.

VII. SYSTEM ARCHITECTURE

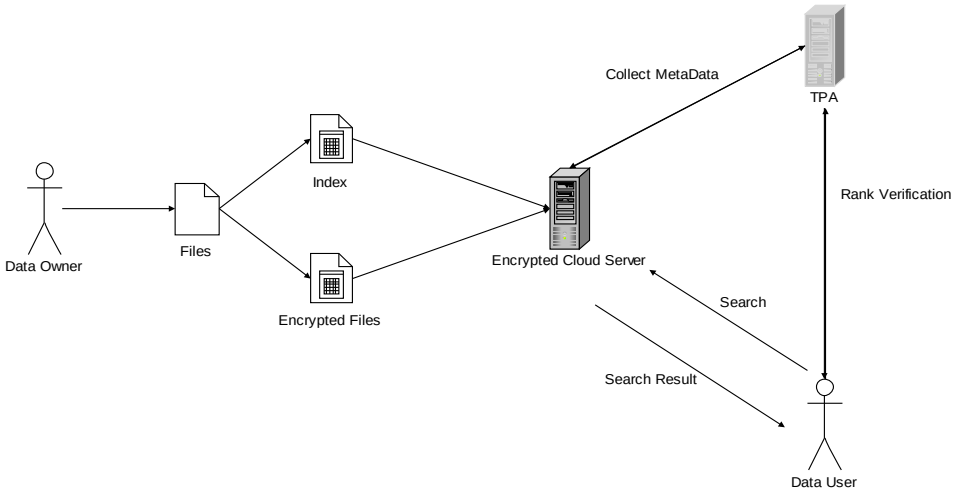

Figure 7.1 system architecture

VIII. CONCLUSION

In this paper, we define and solve the problem of checking the security of the cloud server and data user so that the data user retrieves the correct result from the cloud server by using integrated rank ordered multiple keyword searches over encrypted cloud data. We preserve the privacy of the data and the integrity of rank. By using Hash Mapping technique it solves the challenge of checking the integrity of rank on both cloud server and data user. We can achieve low communication and computation overhead by using an integrated rank order search for effective data retrieval in multiple keyword queries over an encrypted cloud data. As a result of this proposed methodology, we can achieve high security while retrieving the data as well as while outsourcing the data.

REFERENCES

[1] R. Curtmola, J.A. Garay, S. Kamara, and R. Ostrovsky, "Searchable Symmetric Encryption: Improved Definitions and Efficient Constructions," Proc. ACM 13th Conf. Computer and Comm. Security (CCS), 2006.
[2] D. Song, D. Wagner, and A. Perrig, "Practical Techniques for Searches on Encrypted Data," Proc. IEEE Symp. Security and Privacy, 2000.
[3] D. Boneh, G. Crescenzo, R. Ostrovsky, and G. Persiano, "Public- Key Encryption with Keyword Search," Proc. Int'l Conf. Theory and Applications of Cryptographic Techniques (Eurocrypt), 2004.
[4] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure Ranked Keyword Search over Encrypted Cloud Data," Proc. IEEE 30th Int'l Conf. Distributed Computing Systems (ICDCS), 2010.
[5] P. Golle, J. Staddon, and B. Waters, "Secure Conjunctive Keyword Search over Encrypted Data," in Proc. of ACNS, 2004, pp. 31-45.
[6] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-Preserving Multi keyword Ranked Search over Encrypted Cloud Data," Proc. IEEE INFOCOM, 2011
[7] Wang C, Cao N, Li J, Ren K, Lou W **(2010) "**Secure ranked keyword search over encrypted cloud data". In: 30th IEEE International Conference on Distributed Computing Systems (ICDCS). IEEE, Genoa, Italy, pp 253–262

