www.ioirp.com

International Journal of Innovative Research in Computer Science and Engineering (IJIRCSE)
ISSN: 2394-6364, Volume – 1, Issue – 10. September 2015

Secure And Efficient Multi-Keyword Search For Encrypted Cloud Data

MS. S.KALAIVANI
M.E(2nd YEAR)
COMPUTER SCIENCE AND ENGINEERING
RENGANAYAGI VARATHARAJ COLLEGE OF ENGINEERING
SALVARPATTI
kalaisubbaraj@gmail.com

Dr.V.ILLANKUMARAN
PRINCIPAL
RENGANAYAGI VARATHARAJ COLLEGE OF ENGINEERING
SALVARPATTI
v.ilankumaran@gmail.com

***Abstract*—Because of the rising popularity of cloud computing, more and more data owners are motivated to outsource their data to cloud servers for comfort and reduced cost in data management. However, sensitive data should be encrypted before outsourcing for the privacy needs, which obsoletes the use of data such as keyword based search. In existing technique , they widely used on plaintext data to search from cloud server. cannot be directly applied to the encrypted data. In this project we present a secure and efficient multi-keyword search for encrypted cloud data, which concurrently supports the dynamic update operations as the removal and insertion of documents. The vector space model and TF-IDF model are widely combined for the construction of the index and the generation of the request. The AES and secure KNN algorithm is used for encrypting the index and search vectors and at the same time ensure the calculation of accurate relevance score between encrypted index and query vector. To resist statistical attacks ghosts terms are added to the index vector for blinding the search results .**

Keywords—multikeyword, encrypted cloud data, term frequency.

I. INTRODUCTION

Cloud computing is the vision long dreamed of IT as a utility, where the cloud clients can store their distance Data in the cloud to take advantage of the high demand on quality applications and services from a shared pool of configurable IT resources . Its high flexibility and economic savings are motivating individuals and businesses. Local outsource their complex data management system the cloud. To protect the confidentiality of data and unsolicited combat access in the cloud and beyond, sensitive data, such as emails, personal health records, photo albums, tax documents, financial transactions, etc., can be encrypted by data owners before outsourcing the commercial public cloud . This, however, renders obsolete the traditional use of data services based on research by keyword clear. The trivial solution download all data and decryption locally is clearly impractical, because of the enormous amount of bandwidth in cost large-scale cloud systems. In addition, apart from the elimination of local storage management, storage of data in the cloud is useless end unless they can be easily searched and used. So, explore preserving privacy and the efficient search service the encrypted cloud data is of paramount importance. Considering the potentially large number of application data users and huge amount of documents to outsourced data in the cloud, problem is particularly difficult because it is extremely difficult also to meet the performance requirements, system usability and scalability.

PEKS is semantically secure against an adaptive chosen keyword attack polynomial for any striker . The PEKS reveal no information about the message, but allow searching for specific keywords [1].To search over the encrypted data two types of approaches. One possibility is to build an index for every word of interest, lists the documents contain. An alternative is to perform sequential analysis without index. The advantage of using an index is that it can be faster than sequential analysis when documents are large. The disadvantage of using an index is that the storage and updating of the index can be substantial overhead. Therefore, the approach of using an index is more suitable for the data mainly used in read-only [9][10]. There is no tolerance typing errors and minor inconsistencies format, on the other hand, is a typical behavior of the user search and occur most frequently. fuzzy search keywords on effective cloud data encrypted while preserving the intimacy keyword use the edit distance to quantify the similarity keywords and develop a

new technique, namely a wildcard based technique for building fuzzy sets of keywords [4] . The basic idea of LSH is to use a set of hash functions to map objects into several buckets such that similar objects share a bucket with high probability, while dissimilar ones do not and A "Bloom filter" is a bit array that is affiliated with some hash functions. Typographical errors are common both in the search queries and the data sources, but most of the available searchable encryption schemes do not tolerate such errors a fuzzy keyword set based scheme has been proposed to handle the problem [4][6]. Among various multi-keyword semantics, we choose the efficient similarity measure of "coordinate matching". That is as many matches as possible, to capture the relevance of data documents to the search query. Specifically, we use "inner product similarity". propose a basic idea for the MRSE using secure inner product computation, which is adapted from a secure *k*-nearest neighbor (*kNN*) technique [3]. propose an efficient privacy-preserving search method over encrypted cloud data that utilizes *minhash* functions. The fundamental problem of privacy-preserving search is examining the similarity of items. We use a well known technique, known as min hashing to deduce the similarity between sensitive data and the given encrypted query. case there is a need for hiding the access patterns, traditional private information retrieval (PIR) methods or Oblivious RAM can be utilized for the document retrieval process [8]. Boolean search and are not yet sufficient to meet the effective data utilization need that is inherently demanded by large number of users and huge amount of data files in cloud [2]. formally define a secure index and formulate a security model for indexes known as semantic security against adaptive chosen keyword attack (ind-cka) [4].

In this paper present a secure multi-keyword research scheme more encrypted data class clouds, which simultaneously update operations supports dynamic as of deletion and insertion of documents.TF×IDF rule is widely used in the clear from information retrieval, which effectively supports the class multi-search by keyword . The balanced binary tree is widely used to treat the problems of optimization. The keyword balanced binary tree in our system is a dynamic data structure whose node stores a vector . The elements of the vector are the values standardized TF [7][10].

II. RELATED WORK

In this paper , we present a secure multi-keyword research scheme more encrypted data class clouds, which simultaneously update operations supports dynamic as of deletion and insertion of documents. The model of vector space and the TF model of the IDF are combined in the indexes and query construction generation. Model of vector space with TF×IDF rule is widely used in the clear from information retrieval, which effectively supports the class multi-search by keyword . Here the term frequency is the number of times that a given word keyword appears in a document, and the document reverse is obtained through frequency dividing the cardinality of collection of documents by the number of documents containing the keyword . In the model of vector space, each document is identified by a vector, whose elements are the normalized in keywords values TF this document. These elements are the values of the IDF of standardized query keywords in the Document collection. Of course, the lengths of the two vector TF and IDF vector are equal to the total number of keywords, and the scalar product of the vector TF and IDF vector can be calculated to quantify the relevance between the query and the corresponding document.

The balanced binary tree is widely used to treat the problems of optimization. The keyword balanced binary tree in our system is a dynamic data structure whose node stores a vector . The elements of the vector are the values standardized TF. The algorithm secure protective is used to encrypt the index and query vectors, and between-time ensure the accuracy of score calculation of relevance between vectors of index and an encrypted request. In order to withstand the attacks of statistics, Phantom terms are added to the index vector of blinding the search results. To enable the secure, efficient, precise and dynamic multi-password-encrypted key outsourced search for more class data clouds the templates are a dynamic update on the collection of documents , to achieve the effectiveness of sublinear search by exploring an index based on special shaft and a search algorithm efficient and the personal information protection requirements are specified by index Confidentiality and query Confidentiality , trapdoor and keyword confidentiality , unlinkability.

III. ARCHITECTURE DESIGN

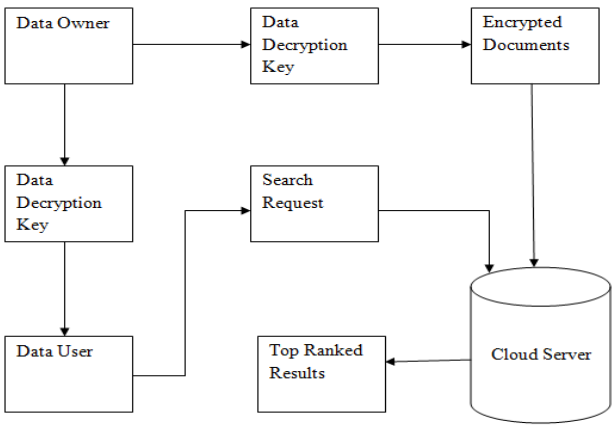

Figure 1. Architecture Design

In this model, the cloud server only knows the encrypted document collection, the searchable index tree, and the search trapdoor *TD* submitted by the authorized user. That is to say, the cloud server can conduct cipher text-only attack. Compared to cipher text model, the cloud server in this stronger model is equipped with more knowledge, such as the term frequency (TF) statistics of the document collection. This statistical information records how many documents are there for each term frequency of a specific keyword in the whole document collection. Equipped with such statistical information, the cloud server can conduct TF statistical attack to deduce or even identify certain keywords through analyzing histogram and value range of the corresponding frequency distributions.

A. Data Owner

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 owner has a collection of documents that he wants to outsource to the cloud server in encrypted form while still keeping the capability to search on them for effective utilization. Data owner firstly builds a secure searchable tree index from document collection, and then generates an encrypted document collection. Afterwards, the data owner outsources the encrypted collection and the secure index to the cloud server, and securely distributes the key information of trapdoor generation and document decryption to the authorized data users.

B. Data Users

Data users are authorized ones to access the documents of data owner. With query keywords, the authorized user can generate a trapdoor *TD* according to search control mechanisms to fetch encrypted documents from cloud server. Then, the data user can decrypt the documents with the shared secret key. The proposed scheme is designed to provide not only multi-keyword query and accurate result ranking, but also dynamic update on document collections. The scheme aims to achieve sublinear search efficiency by exploring a special tree-based index and an efficient search algorithm. The scheme is designed to prevent the cloud server from learning additional information about the document collection, the index tree,
and the query.

C. Cloud Server

The cloud server is mainly considered as “honest-but-curious”, which is employed by lots of works on secure cloud data search. It stores the encrypted document collection and the encrypted searchable tree index for data owner. Upon receiving the trapdoor *TD* from the data user, the cloud server executes search over the index tree, and finally returns the corresponding collection of top ranked encrypted documents. Besides, upon receiving the update information from the data owner.

The advantage of using an index is that it can be faster than sequential analysis when documents are large. The disadvantage of using an index is that the storage and updating of the index can be substantial overhead. Therefore, the approach of using an index is more suitable for the data mainly used in read-only. The basic plan provides provable if the secret pseudorandom function and pseudorandom generator are secure. Controlled research is nearly as good as the basic plan if the primitives are secure. This regime does not support search queries hidden Alice should simply pre-encrypt the plaintext each word separately using a deterministic encryption algorithm. So this stage of pre-encryption as encryption word document with some block cipher.

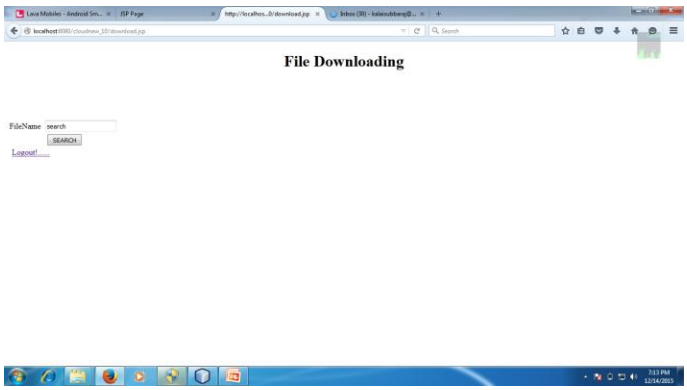

Figure 2. Multikeyword Search

Existing search system will provide the result only based on the Boolean keyword matching system, it means weather it will find the exactly file name same as the keyword than the file will retrieved from the server, it does not provide any search result for misspelled keywords. And also the existing search system never provide the result based on similar keyword .To overcome the above drawback the efficient search system to search the files from the cloud server using multi-keyword. A server to generate the fuzzy keyword set from the file name using the fuzzy keyword set it will create the all possible misspell keywords.

This server will generate three type of keyword set that is K-gram, Wild card and Semantic search. The k-gram have two keyword sets based on edit distance are Edit distance – 1 and Edit distance – 2 . Wildcard have three keyword sets

based on edit distance are insertion, subtraction and deletion for example Keyword is "Mining".The edit distance insert is based on *mining, m*ining, subtraction is *ining, m*ning and the deletion ining, mning and then edit distance two is related on two letters insert, subtract and delete.

Finally encrypt the keyword set and save into cloud server . The semantic search is the synonym based search. Search keyword get encrypt and it will check with the collection of original encrypted file name in the cloud server if the keyword will get matched then connect the fuzzy keyword set for that particular keyword and doing to search the file list based on the Search keyword query, Encrypt keyword and Index based search. Then retrieve the files from the cloud server and consider the searching performance also.

IV CONCLUSION

In this paper, a secure, efficient and dynamic search system is proposed, which supports not only the precise multi-keyword search classified but dynamic insertion and removal of documents. Building a special keyword balanced binary tree as the index, and propose a "depth-first search Greedy" algorithm for better efficiency than linear search. In addition, the parallel search process may be performed to further reduce the cost of time. Presented search system will provide the result only based on the Boolean keyword matching system, it means weather it will find the exactly file name same as the keyword than the file will retrieved from the server, it does not provide any search result for misspelled keywords. And also the existing search system never provide the result based on similar keyword.

To provide the efficient search system to search the files from the cloud server using multi-keyword. A server to generate the fuzzy keyword set from the file name. Here using the fuzzy keyword set it will create the all possible misspell keywords. Search keyword get encrypt and it will check with the collection of original encrypted the file name in the cloud server then retrieve the files from the cloud server and consider the searching performance also.

References

[1] Boneh D., Crescenzo Di.G., Ostrovsky R. and Persiano G. (2004) 'Public Key Encryption with Keyword Search', in Advances Cryptology – Eurocrypt Springer , pp. 506-522G.

[2] Cao N., Lou W., Ren K. and Wang C. (2012) 'Enabling Secure and Efficient ranked Keyword Search over Outsourced Cloud data', Parallel and Distributed Systems, IEEE Transactions on Vol.23, no.28, pp.1467-1479.J.

[3] Cao N., Li M., Lou W., Ren K. and Wang C. (2011) 'Privacy Preserving Multi-Key word ranked Search over Encrypted Cloud data', in IEEE INFOCOM , pp.829-837.

[4] Cao N., Li J., Ren K., Wang C., and Wang Q. (2010) 'Fuzzy Keyword Search over encrypted data in Cloud Computing', in INFOCOM IEEE Procedings, pp 1-5.

[5] Gohetal D. (2003) 'Secure Indexes' , IACR Cryptology eprint Archive , Vol. p.216.

[6] Islam M.S, Kantarcioglu M. and Kuzu M. (2012) 'Efficient Similarity Search over encrypted data', in Data Engineering ICDE , 28th international Conference on IEEE, pp.1156-1167.

[7] Kamara S. and Lauter K. (2010) 'Cryptographic Cloud Storage' in Financial Cryptography and data Security Springer, pp.136-149R.

[8] Kantarcioglu M., Orencik C. and Savas E. (2013) 'APractical and Secure Multi-Key word Search Method over encrypted Cloud Data', in Cloud Computing Sixth International Conference on IEEE , PP.390-397.

[9] Perrig A., Song X.D. and Wagner D. (2000) 'Practical Techniques for Searches on Encrypted Data', in Security and Privacy in IEEE Proceedings, pp.44-55.

[10] Sun X., Wang Q., Wang X. and Xia Z. (2015) 'A Secure and Dynamic Multi-Keyword Ranked Search Scheme over Encrypted Cloud Data', IEEE Transactions on Parallel and Distributed Systems , Vol 1 pp.