



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

An Overview on Fuzzy search Technique for Cloud Encrypted Data

Swara Saoji, Ravi Rajbhure, Nitin Bhil

PG Student, Dept. of CSE., Amravati University, Chikhli, Maharashtra, India

Assistant Professor, Dept. of CSE., Amravati University, Chikhli, Maharashtra, India

Assistant Professor, Dept. of CSE, Amravati University, Chikhli, Maharashtra, India

ABSTRACT: Now a day's cloud computing is one of the most effective data sharing scenario. It is a new model of enterprise IT infrastructure, which can organize huge resource of computing, storage and convenient and on-demand network access to a shared pool of configurable computing resources with great efficiency and minimum economic overhead. Many organizations and individuals are interested in storing their sensitive data eg; personal health record; financial record in cloud. Cloud computing enables the paradigm of data service outsourcing. To protect data privacy, sensitivity cloud data has to be encrypted before outsourced to the commercial public cloud, which makes effective data utilization service and increase accuracy. Fuzzy search technique is used for searching the documents stored on cloud. User will be able to search any documents with the help of keywords. The multi-keyword fuzzy search scheme support more spelling mistakes. In our proposed project, at the time of document searching we will filter the documents from result-set with the help of specified access permission. There is large number of users and huge amount of data files in cloud. Fuzzy search techniques allow users to securely search over encrypted data through keywords. In this paper, we define and solve the problem of secure keyword search over encrypted cloud data.

KEYWORDS: Cloud computing, Fuzzy search, data service outsourcing, IT Infrastructure.

I. INTRODUCTION

Cloud computing is increasingly growing technology which provides an on-demand software, hardware, infrastructure and data storage as services and network computing service. This technology is used worldwide to improve the business infrastructure and performance. The availability of various services over internet is possible through cloud technology which connects software, hardware, data storage and infrastructure. Cloud computing service provider delivers the applications via internet. A public cloud is one based on the standard cloud computing model, in which a service provider makes resources, such as applications and storage, available to the general public over the Internet. Their are various number of users access the Information from Public cloud. Then the security and Authentication of the user is necessary[2].End users can outsource their personal data onto public clouds, and then access those data at anytime and anywhere. Outsourcing data services to the cloud allows organizations to enjoy not only financial savings, but also simplified local IT management since cloud infrastructures are physically hosted and maintained by the cloud providers.To minimize the hazard of data leakage to the cloud service providers, data owners can encrypt their sensitive data, e.g., health records, monetary transactions, before outsourcing to the cloud, while retaining the decryption keys to themselves and other allowed users[2]. Cloud service providers (CSPs) usually impose users' data security through method like firewalls and virtualization but these method do not protect users' privacy from the CSP itself since the CSP possesses full control of the system hardware and lower levels of software stack. There may exist disloyal or curious employees that can access users' sensitive information for illegal purposes[3]. Then the data will be encrypted before outsource the data on public cloud.User can search documents among an encrypted data set stored in the cloud, user have to download it and decrypt the entire data set. Instead of a word-by-word linear scan in the full text search early works built various types of secure index and corresponding index-based keyword matching algorithms to improve search efficiency. All these works only support the search of single keyword though, they support only exact keyword matching but the single-keyword queries are too restrictive for practical use. Misspelled keywords in the

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

single keyword search query gives wrong result or no matching. To overcome this single keyword search, the privacy-preserving multi-keyword fuzzy search over encrypted cloud data is introduced [3].

II. PROBLEM FORMULATION

1. System model:- The system model considered in this paper consists of three entities: the data owner, the data user, and the cloud server. To outsource a set of files to the cloud, the data owner makes a secure searchable index for the file set and then uploads the encrypted files, together with the secure index, to the cloud server. To search over the encrypted files, an authorized user first obtains the trapdoor. A trapdoor is a secret entry point into a program that allows someone that is aware of the trapdoor to gain access without going through the usual security access procedures [1][2]. The data owner outsources the massive size of document to the cloud server with its encrypted data and encrypted searchable index then submits the trapdoor to the cloud server [3]. After receiving the trapdoor, the cloud server executes the search algorithm over the secure indexes and returns the matched files to the user as the search result [1,2]. An additional feature is that the data user may not want to receive all the related documents. Instead, the data user may send a search parameter k along with the search query Q such that the cloud server only returns the top- k most relevant documents. We assume that the data user has the mutual authentication capability with the data owner [3].

Notations :-

- F : the set of original files, assume there are m files. F is denoted as $F = (F_1, F_2, F_3 \dots F_m)$
- C : the set of encrypted files, corresponding to the files in F . Denoted as $C = (C_1, C_2, C_3 \dots C_m)$
- W : keyword dictionary, assume we have n keywords. W is denoted as $W = (W_1, W_2, W_3 \dots W_n)$
- F_{idx} : the keyword set of each file, it is denoted as $F_{idx} = (F_{idx1}, F_{idx2}, F_{idx3} \dots F_{idxn})$
- p : the index vectors for F_{idx} , p is denoted as $p = (p_1, p_2, p_3 \dots p_n)$
- I : the encrypted index vectors for p . I is denoted as $I = (I_1, I_2, I_3 \dots I_n)$
- W_q : a plain text query, assume it contains k keywords, and can be represented as $W_{kw1}, kw2, \dots, kwk$
- q : for a query W_q , the corresponding query vector.
- T , the trapdoor for a query W_q , which is based on q .
- R the list of files in the returned matching result set. It is a sorted list, the order of the files is determined by the scores [5].

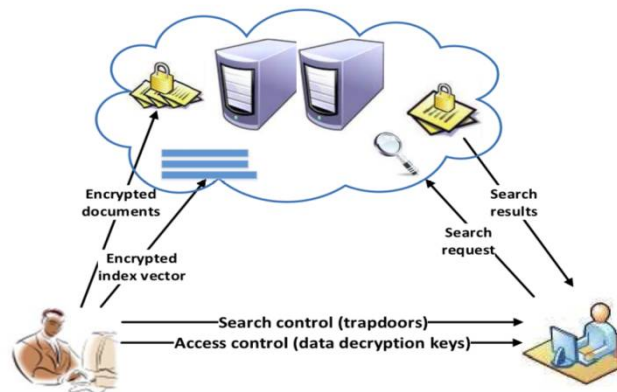


Fig. 1. The system model of our scheme.

• Threat model:-

We assume that both data owners and data users are trusted. But the cloud server is honest-but-curious. Even though data files are encrypted, the cloud server may try to obtain other sensitive information from user search requests while performing keyword-based search over C . So the search should be performed in a secure manner that allows data files to be securely retrieved while revealing as little information as possible to the cloud. Depending on the available information to the cloud server, two threat models are considered here [3].



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

- Know Cipher text Model:-

The cloud server can only know the encrypted files C, the searchable index, encrypted index vector I and the submitted trapdoors T. The cloud server can also know and record the search results. The semantic meaning of this threat scenario is captured by the non-adaptive attack model. We intend to protect the plaintext query/index information against the cloud server and keep the dictionary secret that was used to build the searchable index tree I [1],[2],[3],[4].

- Known Background Model:-

The cloud server knows additional background information in this model. The background refers to the information which can be learned from a comparable dataset. For example, the correlation relationship of two given trapdoors. The main objective of this system is to preserve user data privacy, which includes: 1) file content privacy; 2) index privacy and 3) user query privacy. While file content privacy can be achieved by encryption-before-outsourcing schemes. 4) Keyword privacy: By the search result, the cloud server should not deduce any keyword information of the file set from secure indexes and trapdoors. Keyword privacy requires indexes and queries be properly represented and securely encrypted. 5) Trapdoor unlinkability The cloud server should not be able to link one trapdoor to another even if they are for the same query. Trapdoor unlinkability requires a non-deterministic trapdoor generation function [2],[3],[4].

III. DESIGN GOAL

- User will be able to search any document with the help of keywords Support more spelling mistakes:
- Our multi-keyword fuzzy search scheme should support more spelling mistakes. For example, “network security” related files should be found for a misspelled query “netward security”, “network security”, “network security” and “netwrk security”.
- Privacy guarantee:-The cloud server should be prevented from obtaining additional information from the encrypted data files and the index.
- No pre-defined Dictionary:- No pre-defined dictionary is a great contribution of original scheme, so our scheme should not have pre-defined dictionary.
- Support updating: -The same as original scheme, our scheme should support dataset updating, such as file adding, file deleting and file modifying.
- Ranked results according to the relevance score:- To make users more satisfied with search results, the return results should be ranked according to relevance score.
- Efficiency and Accuracy: -The efficiency of our scheme should be same as the original scheme. And our scheme should be as accurate as possible and keep high accuracy[1],[3],[4]

IV. SYSTEM DESIGN

In MRSE (multi keyword rank search encrypted) technique all keywords are defined in a dictionary and a keyword is identified by its location in the dictionary. Two randomly generated invertible matrices are used for data and file index encryptions. It uses the inner product of two vectors to build the trapdoor for secure keyword queries. It applies an internal ranking algorithm to determine the top k files to be returned to the data consumer. But this method suffers from three major drawbacks. First, it uses a static dictionary[1],[5]. If new keywords are added, the dictionary has to be rebuilt completely. Second, using its trapdoor generation algorithm, an out-of-order problem occurs. Such a problem results in that files with more matching keywords are likely excluded from the top k positions in the matching set. This means the data consumer may not be able to find the most applicable files. Third, MRSE does not consider the effects of keyword access frequencies, thus the files which contain frequent keywords might not be included in the top k return result[5]. To overcome the drawbacks of the MRSE strategy MKQE technique is used. The amount of data is increase from time to time. Thus, the keyword dictionary has to be extended periodically. We propose a new dictionary construction model, introduce a new trapdoor generation algorithm and take the keyword access frequencies into consideration to reduce the query latencies and produce better matching result sets. The keyword dictionary can be prolonged dynamically without touching the contents in the original dictionary. With this strategy, we can greatly reduce the overhead of the dictionary reconstruction as new keywords are added[1],[5]. We design a novel trapdoor generation algorithm. It can effectively reduce the impacts of dummy keywords on the ranking scores. The files which contain more frequently accessed keywords have higher chances to appear at the first k locations of the returned results

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

than the files with less frequent accessed keywords[2.]We formulate the privacy preserving problem of multiple keyword fuzzy search over encrypted data in this section. We denote a keyword collection of a document as an index and an encrypted index as secure index. Similarly, a query is a keyword collection of a search and a trapdoor is an encrypted version of a query. There are three kinds of users Data owner, Data consumer and Cloud Service Provider (CSP).The data owner encrypted the data file before outsource on public cloud by using AES (Advanced Encryption Algorithm) or DES. To design a secure and well performance search scheme over encrypted data, one has to make three important design choices 1). data structure used to build secure indexes and trapdoors; 2). Effective search algorithm that can quantify the level of match between keywords in the query and keywords in the index with high efficiency, and 3). security and privacy mechanisms that can be integrated in the above two design choices thus the index privacy and search privacy can be protected. [2]. It also specifies the set of keywords to form a keyword dictionary to be usedfor queries. we assume the keyword dictionary is dynamic, the DO may add keywords later depending on the changes in the set of sensitive files. For each file, an index vector is generated based on which keywords are contain in it.

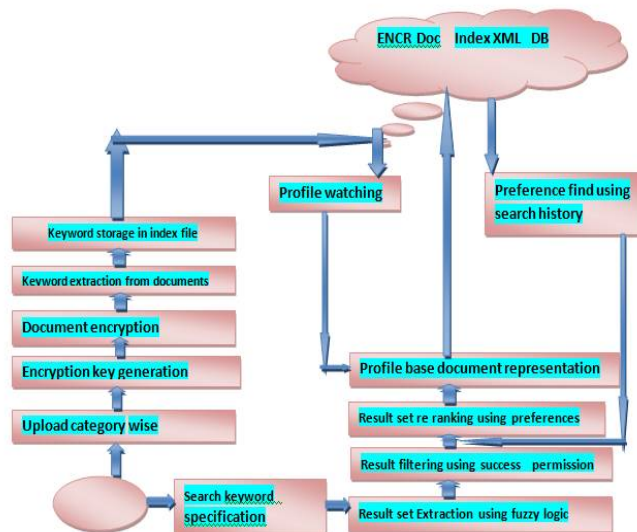


Fig2:-keyword Search engine

Those index vectors are also encrypted and combined together to generate an encrypted index file. Both the encrypted files and the encrypted index file must be uploaded onto the cloud servers. To help secure queries, the DO needs to define a secret key. The secret key contains two invertible matrices and a bit vector. All the elements in the secret key are randomly generated[2],[5]. The secret key is kept on the DO. After the above processes finish, the CSP has all the sensitive files stored in the encrypted formats, no information leakage occurs. Now, the DCs are able to conduct secure multi-keyword ranked queries on those encrypted files. A query is executed as follows. First, a DC sends the set of keywords it is searching for to the DO. Next, the DO builds the trapdoor T based on this set of keywords using the secret key. T is returned to the DC who submits the request. Finally, the DC sends T to the CSP who stores the encrypted files. A matching process based on T is conducted and a set of encrypted files is identified. All these operations on the CSP are executed on the encrypted data only, there's no plain text information exposure to the CSP. Because the number of files which contains one or more keywords specified in the T could be very large, it results in considerable overhead to return all the results to the DC[1][2][5]. In order to select an adequate set of files, a ranking algorithm is applied on these files based on the relevance scoring. Typically, the DC only has to retrieve the top k most relevant files. It sends the requests to the DO for the decryption keys and then decrypts these files[5].



ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

V.CONCLUSION

In this paper we focus on fuzzy search engine on cloud encrypted data. To improve the cloud security and accuracy the data owner encrypted data before outsourcing. Then the fuzzy search technique is used to search the encrypted data. When the amount of encrypted data increases and more keywords need to be introduced, the searching infrastructure can be naturally expanded with the minimal overhead. 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 also design a new trapdoor generation algorithm, which can solve the out-of-order problem in the returned result set without losing the data security and privacy property.

REFERENCES

- [1].Fu, Zhangjie, et al. "Toward efficient multi-keyword fuzzy search over encrypted outsourced data with accuracy improvement." IEEE Transactions on Information Forensics and Security 11.12 (2016): 2706-2716.
- [2].Cao, Ning, et al. "Privacy-preserving multi-keyword ranked search over encrypted cloud data." IEEE Transactions on parallel and distributed systems 25.1 (2014): 222-233.
- [3]. Sun, Wenhai, et al. "Privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking." Proceedings of the 8th ACM SIGSAC symposium on Information, computer and communications security. ACM, 2013.
- [4] Efficient Multi-Keyword Ranked Query on Encrypted Data in the Cloud
- [5] Xia, Zhihua, et al. "A secure and dynamic multi-keyword ranked search scheme over encrypted cloud data." IEEE Transactions on Parallel and Distributed Systems 27.2 (2016): 340-352.
- [6] Orencik, Cengiz, et al. "Multi-Keyword search over encrypted data with scoring and search pattern obfuscation." International Journal of Information Security 15.3 (2016): 251-269.