International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391

Fuzzy Search Engine for Cloud Encrypted Data

Swara Saoji[1], Nitin Bhil[2]

[1]Amravati University, Computer Sci and Engineering, Chikhli 443201, Maharashtra, India

[2]Computer Science and Engineering, Amravati University, Chikhli443201,dist, Buldana, 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 traditional searchable encryption schemes provide a number of approaches to search on encrypted data, but they all support only exact keyword search. Exact keyword search is unsuitable for cloud storage systems, because it doesn't allow users to make any spelling errors or format inconsistencies, and thereby reduces the system usability. The multi-keyword fuzzy search scheme support more spelling mistakes. In our proposed system, fuzzy keyword searching over encrypted cloud data is discussed but access permission verification for searched data is not discussed. 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***

Keywords: Cloud computing, Fuzzy search, data service outsourcing, IT Infrastructure

1. Introduction

Cloud computing is gradually more growing technology which provides an on-demand software, hardware, infrastructure and data storage as services and network computing service. This technology is used universal to improve the business infrastructure and performance. The convenience 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. There are various number of users access the Information from Public cloud. Then the security and Authentication of the user is required [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 risk of data leakage to the cloud service providers, data owners can encrypt their receptive data, e.g., health records, economic transactions, before outsourcing to the cloud, while retaining the decryption keys to themselves and other allowed users [2]. Cloud service providers (CSPs) usually enforce user's 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 fictitious or curious employees that can access users sensitive information for criminal 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 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 introducing [3].

2. 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 protected 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 trap door is a secret entry point into a program that allows someone that is aware of the trap door to gain access without going through the usual security access procedures [1][2]. The data owner outsources the enormous 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].

2. Notations

1) F:the set of original files, assume there are m files. F is denoted as F =(F1,F2,F3...Fm)
2) C: the set of encrypted files, corresponding to the files in F. Denoted as C =(C1,C2,C3...Cm)
3) W: keyword dictionary, assume we have n keywords. W is denoted as W =(W1,W2,W3...Wn)
4) Fidx: the keyword set of each file, it is denoted as Fidx = (Fidx1,Fidx2,Fidx3...Fidxn)
5) p: the index vectors for Fidx, p is denoted as p = (p1,p2,p3...pn)
6) I: the encrypted index vectors for p. I is denoted as I = (I1,I2,I3...In)
7) Wq: a plain text query, assume it contains k keywords, and can be represented as Wkw1,kw2,...kwk
8) q: for a query Wq, the corresponding query vector.
9) T, the trapdoor for a query Wq, which is based on q.
10) 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].

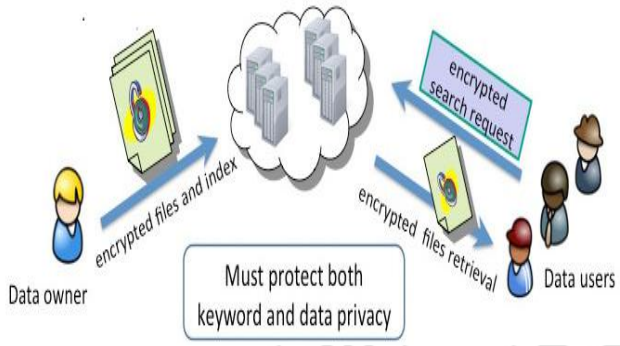

Figure 1: System Model

3. Threat Model

We imagine 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 Cloud. 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].

- **Know Cipher text Model**

The cloud server can only know the encrypted files C, the serchable 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 unlink ability The cloud server should not be able to link one trapdoor to another even if they are for the same query. Trapdoor unlinkability requires an non deterministic trapdoor generation function [2],[3],[4].

4. Design Goals

- 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”, “netwrok security”, “netrwork 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].

5. System Design

There are three kinds of users Academician, User 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. The Data owners upload the data on cloud in encrypted format. The Fuzzy Keyword search Technique is easy way of searching Documents on cloud. It is base on Predefine Keywords set on every Data. The keyword is also encrypted by RSA algorithm. Data owner Uploaded Category wise Data and generating trapdoor key. Trapdoor key is generating for security of Documents and by using Trapdoor Authentication of user is done. The Data consumer searches the document by using keyword. The same trapdoor key is generating for the searchable document for both Data owner and data consumer.AES algorithm is used to encrypt the Document.

The data consumer can download the data and encrypted document is decrypted by DES.

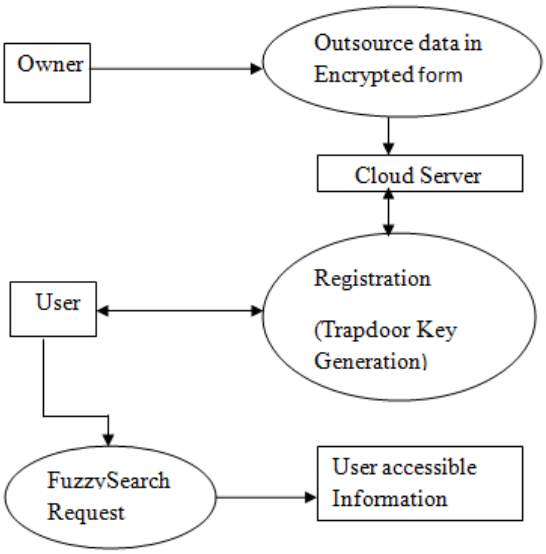

Figure 2: Fuzzy Search Flow

The key idea behind secure fuzzy keyword search is two-fold:

1) Building up fuzzy keyword sets that incorporate not only the exact keywords but also the ones differing slightly due to minor typos, format inconsistencies, etc.;
2) Designing a storage-efficient and secure searching approach for file retrieval based on the resulted fuzzy keyword sets [12].

After constructing the fuzzy keyword sets, the search schemes goes as follows: (i) The data owner first compute the fuzzy keyword set and then he computes Trapdoor set {Twi}, for each word Wi ∈ fuzzy keyword set with a secret key sk, which is shared between data owner and user. The data owner encrypts the file id using secret key and outsourced the index table and encrypted file id to cloud storage.

1) The user then computes the trapdoor {Twi'} for searching in send it to server.
2) Upon receiving this request from user, server compares this trapdoor in index table and returns the matched file ids. Then user decrypts the returned files and retrieves relevant information.

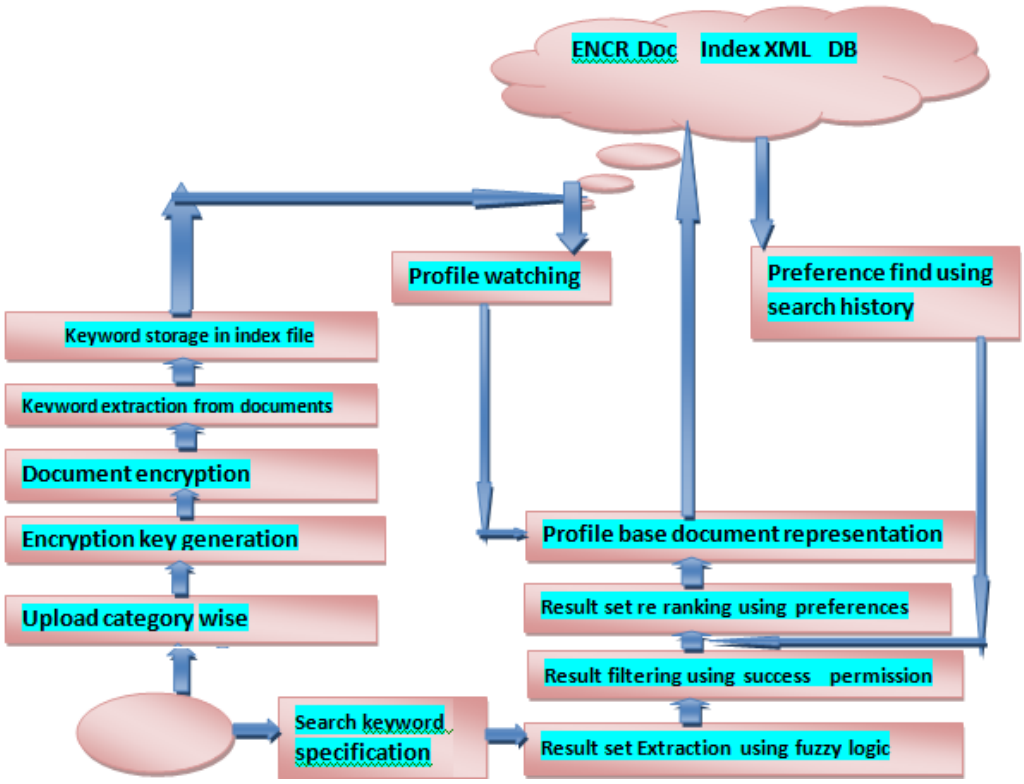

Figure 3: keyword Search engine

6. 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 use 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. It also maintains the privacy of the keyword and the file associated with the keyword. 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. In this paper, we design fuzzy search over cloud and also eliminate the problem of exact match search, increase the possibility of search by allowing typo error up to some extent.

7. Future Scope

Fuzzy search technique for cloud encrypted data is the keyword base search of encrypted data on cloud. The encrypted data on cloud can easily search by Fuzzy search technique and user can download the encrypted data to use. Trapdoor key is used for the Authentication of user as well as Privacy and security purpose**.** In the future scope of this system we are willing to do this application as a mobile aap. Users can use this application on their mobile phone. The list of predefined keyword can also increase for searching documents and vast number of users can connect to it. Then this application can be used in big companies and big organizations.

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.

Author Profile

Swara Saoji received the B.E and M.E. degrees in Computer Sci and Engineering from Amravati University, Maharashtra, India.