A Survey on Efficient Search Schemes over Encrypted Data in Mobile Cloud Environments for Secure Healthcare Data Management and Privacy Preservation

Dr. Hana Kim[1]
Prof. Giulia Conti[1]
[1] University of Zurich, Department of Cybersecurity and Cloud Computing for Biomedical Data Systems, Zurich, Switzerland

ABSTRACT

Cloud computing is computing based on internet. Cloud storage is a service model in which data is maintained, managed, backed up and it is available to users over a network. Cloud storage is massive storage. The privacy issue is a major problem while storing data in the cloud. While storing data in the cloud it doesn't ensure any security. The encryption of a large volume of data is very difficult. So the data owners have as much as trust in storing data in the cloud. To solve this problem encryption is the accurate method. Encrypt the whole data before uploading the file into the cloud. If data user wishes to retrieve any data then retrieve and decrypt the file after downloading that. The retrieval is done by encrypted search scheme. It is the need or importance of encrypted search scheme over the mobile cloud. Higher power consumption for mobile devices for encryption is a major problem. The encryption process is more complicated based on capacity and battery life. Traditionally, there are three categories of search schemes: Single keyword search, Boolean keyword search, and Ranked keyword search. In this paper, provide a survey on different Search Scheme over Encrypted Data on Mobile Cloud. It also includes a survey on various techniques that enhance the Ranker Keyword Search.

KEYWORDS: Search Schemes; Encrypted Search; Ranked Search; Mobile Cloud.

1. INTRODUCTION

Cloud computing, as a promising computing model, enables users to remotely store their data into a cloud. Cloud computing is an emerging technology in recent years. The cloud has several components, such as data owner, data user, service application cloud server etc. There are many benefits in cloud computing such as reduced costs, always-on availability, scalability, flexibility, improved mobility, reliability and so on. Cloud computing eliminates the cost of buying software and hardware. And hence it reduces the cost and capital expenditure. Cloud computing provides always on availability, i.e. the connection is always on. As long as the connection exist the user can get the application. It has improved mobility. Data can be accessed by the data user from anywhere in the world. Cloud computing also provides improved scalability, reliability, and no maintenance is needed. Cloud storage is a massive storage. The privacy issue is a major problem while storing data in the cloud. While storing data in the cloud it doesn't ensure any security. The encryption of a large volume of data is very difficult. So the data owners have as much as trust in storing data in the cloud. To solve this problem encryption is the accurate method. Encrypt the whole data before uploading the file into the cloud. If data user wishes to retrieve any data then retrieve and decrypt the file after downloading that. The retrieval is done by encrypted search scheme. It is the need or importance of encrypted search scheme over the mobile cloud.

Traditionally, there are three categories of search schemes: Single Keyword Search, Boolean Keyword Search, and Ranked Keyword Search. In Single Keyword Search, we can use single keyword query. In which the whole data is divided into words and each word of the document is encrypted. Boolean Keyword Search is based on the boolean operation. There are conjunctive keyword search and disjunctive keyword search. It is the boolean operation, conjunction and disjunction, i.e AND and OR. Ranked Keyword Search is based on relevance score. Here is the retrieval of top-k ranked files. It improves the system efficiency by returning the files in a specific ranked order. It also reduces network traffic and reduces search and retrieval time.

2. RELATEDWORK

In this section, we study various Search Scheme over Encrypted Data on Mobile Cloud such as Single keyword search, Boolean keyword search and Ranked keyword search. We compare these techniques, their performance, efficiency, and computational overhead.

Encrypted Search Scheme

Single Keyword Search: This method [4] was proposed by Song et al. Song et al. They proposed a method, in which the document is divided into words. Each word of this document is encrypted and uploaded into the cloud. It is not as much as efficient as the other two methods. Here, say that each document is divided into a

number of words. Each word may be any token, i.e. it may be an English word, a sentence etc. based on the application domain. Each word is a 64-bit block. The divided words have the same length. When the user has to search for any file, the data owner will send the key to the user if it is an authorized one. So the user can determine each document that contains the word W without learning anything else. The data owner encrypts the document which contains the sequence of words W_1; W_2; :::; W_l. If the user has to search for the word W, then the data owner sends the key k_i to the user. This allows the user to determine each document that contains the word W but doesn't reveal anything where W_i 6= W . Hence say that this is a controlled searching method. That is if the data owner wishes to reveal the key then only the user can determine the document.

There are two types of approaches for searching. One is to build up an index for each word W of the document. The index lists the documents that contain W. Actually it lists the position. The second method is a sequential scan without an index. That is it is the linear search. In which it searches each document one by one and check for the word W. Among both the scheme, an index based search is faster than the sequential search. Its advantage is raised when the document has a large size. Even though indexing is faster, it also has a disadvantage. Managing the index, i.e. storing and updating the index is difficult and it is a problem.

Boolean Keyword Search: Encrypted search scheme includes Boolean keyword search. In Boolean keyword search [4], [7], the server sends the file based on the presence or absence of the keywords. It doesn't bother about the relevance of the document. So the server may send back the least relevant files first based on the presence or absence of the searched keyword. In boolean Keyword Search, there are two search schemes, conjunctive keyword search, and disjunctive keyword search. It is based on the boolean operation conjunction and disjunction, i.e AND and OR. In Conjunctive keyword search, it retrieve the documents which contain all the keywords that are in the query. Or it returns the documents that don't contain those keywords. So, basically, it returns "all-or-nothing". In the conjunctive keyword search the query is in the form of conjunction, i.e the conjunction of the query keywords.

In the disjunctive keyword search, it returns the documents that contain a subset of the specific keywords. In the disjunctive keyword search the query is in the form of disjunction, i.e the disjunction of the keywords in the query. So the retrieved documents contain the subset of the keywords in the query. In Boolean Keyword Search, can use multiple combinations of conjunction and disjunction. For example, "Great ^ (English_Britain), one could send the documents for both "English" and "Britain" and for "Great". In the database, boolean search is the widely used method. Any query which contains an AND, OR or NOT requires boolean support to execute.
Ranked Keyword Search: The ranked encrypted search [5], is more efficient than the others. It is based on the index oriented method. Here is the retrieval of top-k ranked files. The ranking is based on the relevance score. It improves the system efficiency and usability by retrieving the files in a specific ranked order. The ranking is regarding specific criteria, such as keyword frequency. The keyword frequency is simply the term frequency. Term frequency is, the number of times the term "t" that have to be searched occur in the document. It also includes the weight information, such as the relevance score of each file. The score is embedded before uploading the file collection. Wang et al. [3] proposed a ranked keyword search, which proves that Ranked search improves the system efficiency and usability. There is a relevance score computation method,

$$Score(t; F_d) = (1 \div iF_{dj}) \times (1 + Inf_{d;t})$$

Where F_d represents the length of the F_d, $f_{d;t}$ denotes the term frequency(TF) Of term in the file. This equation satisfies the single keyword search.

Various Schemes to Improve Power and Traffic Efficiency
Traditional Search Scheme: This is the earlier search scheme for mobile cloud. Among the three search methods ranked keyword search is most efficient. Traditional search scheme is one of the techniques that enhance the ranked keyword search. It is to improve energy consumption. The data provider first performs the stemming i.e removal of stop words. Then perform indexing work [1]. The data owner first stem the document then encrypts and hashes each term and find the index. Then the data owner also encrypts and stores the index into the cloud server. The file collections are also encrypted and stored into the cloud server

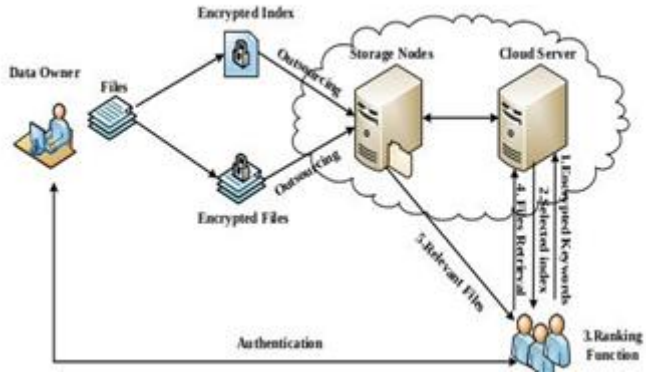

Fig1. Traditional Search Scheme

This is a Two-Round-trip Search(TRS) scheme. It is very complicated while compared to the PlainText Search scheme (PTS). In TRS the file search and retrieval are done in two round whereas in PTS it is done in one round. The traditional search scheme the score calculation and the retrieval of the top-k relevant files and the decryption of the files sent by the cloud server. So it is known as Two-Round-trip Searching scheme. In traditional search schemes, the mobile user has as much as work due to this Two-Round-trip Search scheme. It will increase the energy consumption, network traffic and search and retrieval time. If the user has to search for any document, then the owner first checks whether the user is an authorized one. The user is authenticated with identity and the owner sends the key to the user. The user generates the query and applies stemming and indexing operation on the query. Then encrypt it with the key provided by the owner and hashes it. The keyword in an encrypted fashion is sent to the cloud server. The cloud server receives the keyword in encrypted fashion. Then it searches in the index. Then the server searches for the index related to this keyword and sends the keyword to the user. User have to calculate the relevance score and retrieve the positions. Hence to find the top-k relevant files and sends it to the cloud server. The cloud server selects the files and sends back to the user. The data user decrypts the files and receives the original data.

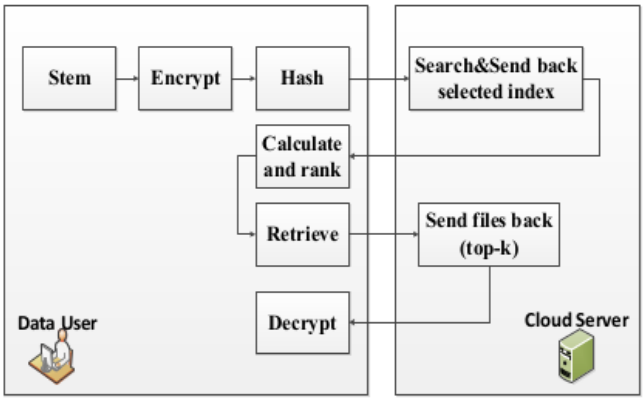

Fig1. ***Two-Round-Trip Encrypted Search***

TEES: Traffic and Energy saving Encrypted Search: The Traditional Search Scheme is an encrypted search scheme with a low security level. So introduce a new architecture for encrypted search scheme. It provides high security, known as TEES(Traffic and Energy saving Encrypted Search) [1]. There is also the process of authentication. First, there is the indexing process then the index is encrypted and stores into the cloud along with the encrypted file search. The data user is authenticated with the data owner by using identity. If a data user wishes to search for any files, then the owner first checks whether the user is an authorized one. The user generates the query and applies stemming and indexing operation on the query. Then encrypt it with the key provided by the owner and hashes it. The data user wraps the encrypted keyword into a tuple by adding some noise and the keyword in encrypted fashion is sent to the cloud server. The wrapping provides an additional level of security. The cloud server calculates the relevance score and retrieves the positions, hence find the topk relevant files and sends it to the data user. It will reduce the burden to calculate relevance score to the user. The data user decrypts the files and receives the original data.

EnDAs: efficient Encrypted DAta search: EnDAs ie the latest technology for search scheme. There is the concept of trapdoor generation (encrypted keyword) and Trapdoor Mapping Table. It will reduce the search delay and network traffic. For trapdoor generation, EnDAS [2] stores a Trapdoor Mapping Table (TMT) in a mobile device. In which stores the trapdoors and corresponding keywords. That is there is the mapping of keywords which we are used and the corresponding trapdoor generated. So it acts as a cache. Here is also the authentication phase. But the encryption of query is done by the data owner. The Trapdoor Mapping Table is maintained at the mobile device side. So, when the mobile device starts a search request, the trapdoor is searched at the table.

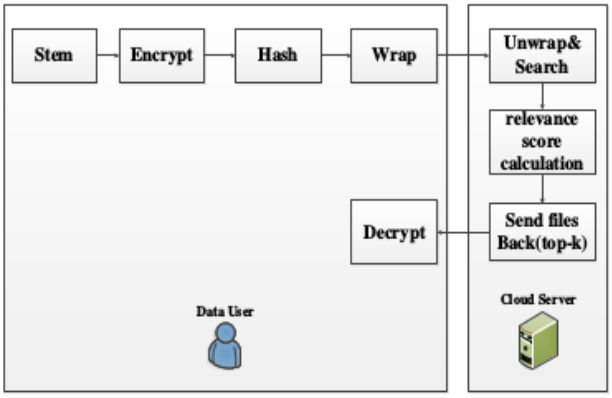

Fig2. One-Round-Trip Encrypted Search

If it is not there then only send the query to the owner and it will generate trapdoor. This is a One-Round-trip-Search scheme because the score calculation and the relevant file retrieval is done by the cloud server. Furthermore, EnDAS also provides compressed and reduce network traffic to transmit trapdoors.

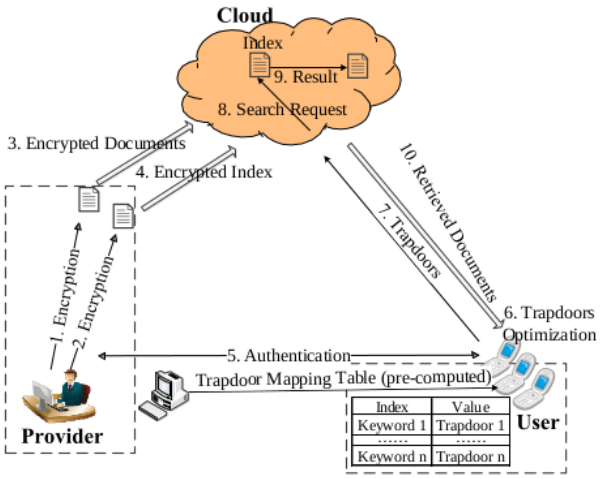

Fig3. EnDAS system over mobile cloud

3. COMPARISON

In this section, we compare these schemes which we survey. we compare the Single keyword search, Boolean keyword search and Ranked keyword search schemes. Single Keyword Search can use single keyword query. In which the whole data is divided into words and encrypt each word. It deals with sequential search, so not get relevant data and it is more time-consuming. Boolean Keyword Search support both conjunctive and disjunctive search. The Ranked Keyword Search is much better than Boolean Keyword Search. In which we get the most relevant data it is less time-consuming. Among the three methods Ranked Keyword Search is better. Among the techniques that enhance Ranked Keyword Search is EnDAs. EnDAs provides compressed and reduces network traffic to transmit trapdoors.

4. CONCLUSION

In this paper, we survey various Search Scheme over Encrypted Data on Mobile Cloud. There are many keyword search methods such as Single keyword search, Boolean key-word search and Ranked keyword search.

We compare these schemes. Cloud storage is massive storage. The privacy issue is a major problem while storing data in the cloud. While storing data in the cloud it doesn't ensure any security. The encryption of a large volume of data is very difficult. So the data owners have as much as trust in storing data in the cloud. To solve this problem encryption is the accurate method. Encrypt the whole data before uploading the file into the cloud. If data user wishes to retrieve any data then retrieve and decrypt the file after downloading that. The retrieval is done by encrypted search scheme. It is the need or importance of encrypted search scheme over mobile cloud. Ranked keyword search is the most efficient keyword search method. Through which we get the most relevant data, and can retrieve top-k relevant data. There are many search schemes based on the ranked search scheme. The modification in which built an efficient scheme. The Encryption DAta Search use trapdoor table. It will reduce the search time and network traffic.

REFERENCES

[1] Jian Li, Ruhui Ma, Haibing Guan, "TEES: An Efficient Search Scheme over Encrypted Data on Mobile Cloud", IEEE Transactions On Cloud Computing, 2015.
[2] Ruhui Ma, Jian Li, Haibing Guan, Mingyuan Xia and Xue Liu, "EnDAS: Efficient Encrypted Data Search as a Mobile Cloud Service", IEEE Transactions On Emerging Topics In Computing, 2015.
[3] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy preserving multi-keyword ranked search over encrypted cloud data", IEEE Transactions on, vol. 25, no. 1, pp. 222233, 2014.
[4] D. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data", 2000 IEEE Symposium on. IEEE, 2000, pp. 4455.
[5] . Wang, N. Cao, K. Ren, and W. Lou, "Enabling secure and efficient ranked keyword search over outsourced cloud data", IEEE Transactions on, vol. 23, no. 8, pp. 14671479, 2012.
[6] A. A. Moffat, T. C. Bell et al., " Managing gigabytes: compressing and indexing documents and images", Morgan Kaufmann Pub, 1999
[7] Y. Chang and M. Mitzenmacher, "Privacy preserving keyword searches on remote encrypted data", in Applied Cryptography and Network Security. Springer, 2005, pp. 391421