International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 06 | June 2019 www.irjet.net p-ISSN: 2395-0072

Multiple Keyword Search over Encrypted Cloud Data

Ms. Shraddha S. Karande, Mr. N.V. Kurhade

[1]Student in Department of Computer Engineering, SPCOE, Otur
[2]Professor in Department of Computer Engineering, SPCOE, Otur

---

Abstract - *With the arrival of cloud, it's clad to be increasingly rife for knowledge homeowners to source their info to open cloud servers whereas allowing info shoppers to recover this information. For protection considerations, secure ventures over encrypted cloud info have impelled many analysis works below the one owner model. In any case, most cloud servers in apply do not merely serve one owner; rather, they support numerous home owner share the benefits brought by distributed computing. Hence these knowledge homeowners will transfer their files over the cloud so the different knowledge users might realize various files and use them.*

***Key Words*: Cloud computing, Multi-keyword, Ranked search, multiple data owner**

1. INTRODUCTION

In today's world, the sharing of data is a vital part of our lives. The data can be sensitive and hence when it gets into the hands of wrong people, it can turn out to be harmful for both the owners and the receivers. Hence to overcome such a situation, encryption and decryption of data can be done for the safe exchange of any amount and type of data. There are two sets of parties which would use this technique; one being the data owners, which would own the data and the next are the data users which do not own the data but use this after getting the permission for the usage. The data owners and the data users need to be authenticated on the cloud servers beforehand to use any of the services provided. When the sensitive data is outsourced to the cloud, so as to enable the easier accessing of the data by the data owners and the data users, it is encrypted. The data encrypted has a list of keywords which are sent to an administration server. This in turn is then re-encrypted and uploaded by the administration server.

When the data users would want to access these encrypted files, they will have to get themselves authenticated. Once the data users are authenticated and verified, they would search the files using keywords. The keywords are sent to the administration server which in turn would encrypt the given keyword. The encrypted keyword is then compared to the existing keywords and the files are given to the data users after the decryption. Hence this helps in creating a secure environment for the exchange of the information among the data owners and the data users.

1.1 Multi-Keyword

Multi-keyword refers to the ability of searching multiple keywords at a given time. This means that the user can search n number of keywords at a given time. The file retrieved is the one which either contains all the keywords or a minimum of one of the keyword that the user searches.

1.2 *Ranked Search*

As the user searches for the document over the cloud, like Google, the cloud server can return the document which is the most relevant document amongst the entire collection of document. The search result can be ranked on different parameters. The parameters can be last downloaded, last visited, recently uploaded, etc. when the user tries to access a document, he/she can specify the k number of documents to be downloaded. When the k is specified, the top k documents are downloaded. This saves a round trip of requests, bandwidth; as only specified numbers of documents are downloaded and the computation time is reduced.

2. Literature Survey

Michael Armbrust *et al* [2] explain about the various advantages and uses of the cloud storage. It helps in a large amount of storage space without the use of the resources in the real time. The resources used to carry out storage operations are never owned by the users and hence the users have to pay per use. This strategy helps in eco friendly and green computing as well. This is possible as the resources that were used earlier for the storage, such as the servers, systems, space, cooling systems and many more are no longer required. The storage systems are virtually present for the user and logically present in a different location. The paper also discusses about the different service models; namely, Infrastructure as a service, Platform as a service

and Software as a service. It also discusses the different type of clouds such as public cloud, private cloud, hybrid cloud and the community cloud.

Dawn Xiaodong Song *et al* [3] discusses about the four of the most important concepts of provable security, query isolation, controlled searching and hidden query. The provable security helps in keeping the data secure. This is done as the untrusted server cannot understand about the plaintext from the encrypted text that is uploaded. The query isolation also helps in maintaining the secrecy.

The untrusted server will be unable to learn anything about the plaintext present in the file through the encrypted data about which it is queried. The controlled searching explains that the untrusted server will be unable to search any query or a keyword without the authentication of the users which are already registered. The hidden query talks about the keyword or the query that is being searched. This is in a non-readable form and therefore the untrusted server does not know what the user is searching. Hence the untrusted server will never be able to guess or hack into the system and retrieve any file saved.

Cong Wang *et al* [4] discusses over the drawbacks of the traditional cloud file retrieval system. There are majorly two drawbacks of this system. The user when tries to retrieve a file from the cloud, the user has to download all the files that are related to the keyword or the query that the user has entered. This leads to a huge consumption of the bandwidth. As the user has downloaded a large number of files, user has to decrypt each file in order to understand whether the file is required or not. The file that is retrieved can be either useful to the user or can be an older file which has not much significance.

Therefore, the authors put forward an idea about ranking of the files and documents that are saved over the cloud. This would help in retrieving the files which are recent or most downloaded. The paper also discusses the shortcomings of the Searchable Symmetric Encryption (SSE).

Jin Li *et al* [5] discusses about the fuzzy keywords that might occur during a search. Fuzzy keywords are the words which are spelled wrong and are entered into the query box. They can be done using the simple spell checking. These kinds if words should be given suggestion and have to show results which might be possibly near the word. For example, if a query is made for Lonfon, the fuzzy keyword correction should show any data that might contain Lonfon or London as it is the nearest word which might be correct. It may also consider every possible keyword. This might include *Lonfon, L*onfon, etc.

Qin Liu *et al* [6] discusses about the ADL. The aggregation and distribution layer (ADL) is a middleware layer between the users and the cloud. It was envisioned such that an ADL will be deployed in an organization that has outsourced the data operations to a cloud. The ADL will aggregate queries from multiple users and send a combined query to the cloud. Due to this combined query, the cloud will need to execute the query only once and return all matched files to the ADL. Furthermore, since the files of most interest to the users need to be returned only once, the communication costs will also be reduced.

Furthermore, the authors have discussed over the Efficient Information retrieval for Ranked Query (EIRQ). The EIRQ helps in the retrieval of the files which are the top k specified. The user has to define the value of the k percentage of the files that are required by the user. This will help in lowering the overhead communication cost and hence the computations required will be less. This would save a round trip time and hence would give better and faster results.

The author Hongwei Li *et al* [7] describes the arrangement of a model which involves a single data owner, multiple data users and a single cloud server. This model is very similar to the multi-owner multiple data users model. As the data owner is single, the key used to encrypt the files is same.

This method of uploading the files is a disadvantage. If there is a leakage of key, the cloud server along with the other users will be able to retrieve the files that are stored in the cloud server. If this happens, a lot of sensitive information can be compromised.

Wei Zhang *et al* [8] discusses about extending the existed single-owner scheme to a full-fledged multi-owner scheme will cause abundant problems. In the single-owner scheme, once a data user wants to issue a keyword search, he has to ask the data owner for secret keys to generate trapdoors (encrypted keywords). Unfortunately, when there are multiple data owners, asking different data owners for keys to generate trapdoors would be infeasible.

First, not all data owners are always online simultaneously when a data user wants to perform a query. If data owners are offline, these owners' data can't be retrieved in time. Second, in order to search different owners' data, data user has to generate a specific trapdoor for each data owner, sending these trapdoors to the cloud server would cause considerable communication overhead. An alternative solution is to share a secret key among all data owners. However, this measure will lead to the security threat of single point of failure.

PROPOSED SYSTEM

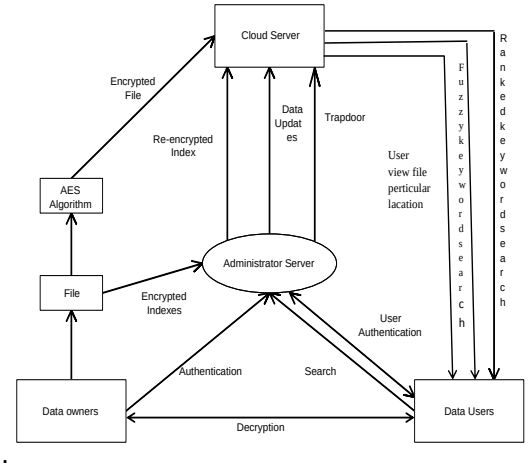

Fig. 1 : System architecture

The system consists of four components: the data user, data owner, administrator and the cloud server. The data owners are the group of users which own the data and upload their data over the cloud server. The data users are the users which try to access the uploaded data. The administrator acts as a mediator between the data owner or data user and the cloud server. The cloud server is the huge collection of data which is made available for all the data owners to upload files and documents and also for the data users for the successful retrieval of the data.

The data users and data owners have to authenticate beforehand. This will help in preventing any unauthorized trespassing. The administrator maintains a system for publishing the trapdoor whenever the data user asks for a specific file or folder. The trapdoor is the encrypted keyword which is to be searched over the cloud server. Once this is sent, it will help in safe retrieval of the documents without the cloud server getting to know about the plaintext.

After the files are found, the cloud server sends the encrypted files to the respective data user. The data user will have to decrypt the file with a key and access the contents in it. user can search any fuzzy keyword from a perticular location only.We also show the ranked keyword search by the user.Finally we show the count of user search file and actually file downloaded by the user.

Advantages:

1. Multi-keyword ranked search over encrypted cloud data (MRSE)
2. Providing security to data with advanced function.

The proposed scheme ensures that only authenticated data users can perform correct searches. Moreover, once a data user is revoked, he can no longer perform correct searches over the encrypted cloud data.

3. CONCLUSION

A thorough analysis of different methods of storing data, the formation of the dictionary, maintenance of the dictionary, the ranking of the files that are to be retrieved, the number of files to be retrieved was done. The above mentioned methods have helped in a better understanding. A secure method of retrieving data from cloud would help the users to trust the cloud services, as it is not very widely used due to a misconception of lack security. As more and more data would be saved over the cloud, the data can be retrieved at any given point in time and place.