



# Secure Outsourcing , Updation And Retrieval Of Data Over Cloud

Haseena.M.K

Assistant Professor

Department Of Artificial Intelligence & Cyber security  
Illahia College Of Engineering and Technology, Muvattupuzha, India

**Abstract:** Cloud computing has emerging as a promising pattern for data outsourcing and high-quality data services. However, the sensitive information on cloud potentially causes security problems. Encryption techniques protects data security to an extent, but at the cost of compromised efficiency. In this paper, we focus on secure updation of data in cloud. The privacy issue is formulated from the aspect of similarity relevance and scheme robustness. Leakage of data is eliminated through a two-round searchable encryption (TRSE) scheme .All the computing work is done on the server side by operations only on cipher text.

**Index Terms** - Cloud, Encryption, Ranking, Vector space, Cryptography.

## I. INTRODUCTION

Cloud computing a critical pattern for advanced data service However, security have been incessantly presented as outsourcing of sensitive information including health history, photos and emails is spontaneously expanding. Data loss and privacy breaches reports in cloud computing systems appear from time to time. Data privacy threats roots in the cloud itself . The cloud service providers are able to control and monitor the data and the communication between users and the cloud at will, lawfully or unlawfully, when users outsource their private data onto the cloud.

### 1.1 Privacy In Cloud

In cloud computing, data owners may share their outsourced data with many users, who want to only retrieve the data files they are interested in keyword-based retrieval is a popular retrieval method but it turns out to be a difficult task in cipher text scenario due to limited operations on encrypted data. It is preferred to get the retrieval result with the most relevant files that match with users interested files only and so the files should be ranked in the order of relevance by users interest and only the files with the highest relevances are sent back to users. Preventing the cloud from involving in ranking and entrusting all the work to the user is a natural way to avoid information loss. But the limited computational power on the user side and the high computational overhead precludes security of information. The main threat on data privacy roots in the cloud itself . To ensure privacy, users usually encrypt the data before outsourcing the data to cloud. One of the most popular ways to do so is through keyword-based retrieval. It is a typical data service in plaintext forms in which users retrieve relevant files in a file set based on keywords and it turns out to be a difficult task.

A series of searchable symmetric encryption (SSE) schemes have been proposed to enable search on cipher text. Traditional SSE schemes support only Boolean in cipher text scenario due to limited operations on encrypted data keyword search, without considering the difference of relevance with the queried keyword of these files in the result. In the former, files are ranked only by the number of retrieved keywords, which impairs search accuracy. Preventing the cloud from involving in ranking and entrusting all the work to the user is a natural way to avoid information loss. The secure multi keyword top-k retrieval over encrypted cloud data is provided by making the cloud do more work during the process of retrieval without information leakage.

## II. LITERATURE REVIEW

### 2.1 . Traditional Searchable Encryption

It focus on security definitions and efficiency of encryption, support only Boolean keyword retrieval without ranking. It is a traditional searchable encryption schemes allow a user to securely search encrypted data over cloud. Here typical user searching behavior happen very frequently.

### 2.2. Secure Rank-Ordered Retrieval With Improved Searchable Encryption

Zerr et al. proposed a ranking model to guarantee privacy-preserving document exchange among the groups, which allows for privacy-preserving top-k retrieval from an outsourced index. They proposed a relevance score transformation function to make relevance scores of different terms indistinguishable and such that improves the security of the indexed data.

### 2.3. Top-k Retrieval Over Encrypted Data In Cloud

On the basis of SSE, they proposed the one-to-many OPM to further improve the efficiency .However, these schemes support only single keyword retrieval. When large number of data users and documents in the cloud, multi keyword in the search request is considered and return the most relevant documents in the order of their relevancy with these keywords.

### 2.4. Top-K Multi Keyword Retrieval Over Encrypted Cloud

They employed coordinate matching and inner product similarity to measure and evaluate the relevance scoring. Here homomorphism is used to preserve the data security. They devised a secure protocol for processing k-nearest-neighbor index query both the data security of the owner and the query privacy of the client. Boolean representation in their searchable index is employed in two schemes, 1 denotes the corresponding term exists in the file and otherwise zero. Since all these server-side schemes employ server-side ranking based on OPE.

## 2.5 Towards Secure Multi keyword Top K Retrieval Over Cloud

We consider a cloud computing system hosting data service, in which three different entities are involved. They are cloud server, data user and owner. The cloud servers cannot be fully entrusted in protecting data, since data may contain sensitive information. Outsourced files must be encrypted for this reason. Any kind of information loss that would affect data security are regarded as unacceptable. To outsource onto the cloud server in encrypted form the data owner has a collection of  $n$  files to and expects the cloud server to provide keyword retrieval service to data owner himself or other users. The data owner needs to build a searchable index  $I$  from a collection of  $l$  keywords extracted out of  $C$  to achieve this, and then outsources both the encrypted index and encrypted files onto the cloud server. The user side computing power is limited, which means that operations on the user side should be simplified. Thus, the data user encrypts the query and sends it to the cloud server that returns to the data user the relevant files. The data user can decrypt and make use of the files afterwards. In this paper, we introduce the concepts of similarity relevance and scheme robustness to formulate the privacy issue in searchable encryption schemes, so the insecurity problem is solved by proposing a two-round searchable encryption (TRSE) scheme. Homomorphic encryption and vector space model are employed. In the proposed scheme, the majority of computing work is done on the , which guarantees top-k multi keyword retrieval over encrypted cloud data with high security and practical efficiency.

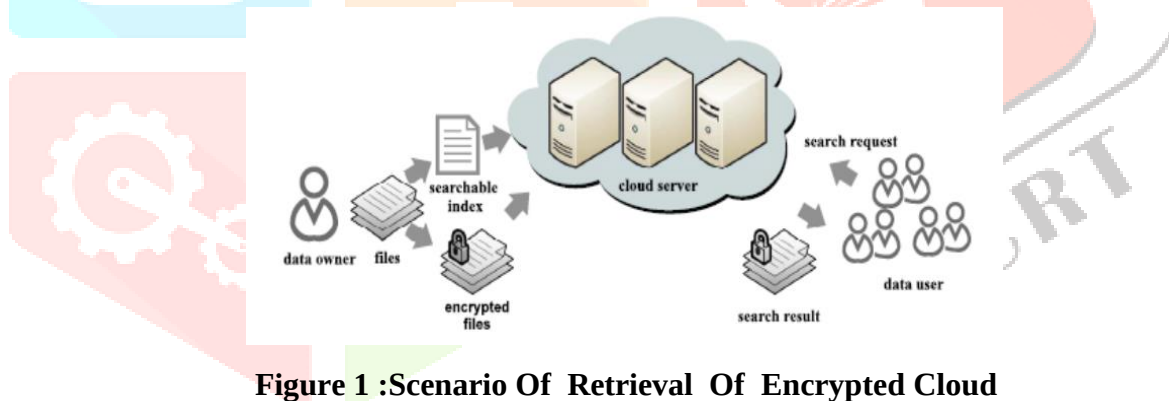


**Figure 1 :Scenario Of Retrieval Of Encrypted Cloud**

### Relevance Scoring

Some of the multi keyword SSE schemes support only Boolean queries, that is a query matches with a file. Considering the large number of data users and documents in the cloud, we have to consider multi keyword in the search query and return documents in the order of their relevancy with the queried keywords. To weight the relevance, scoring is a natural way. Based on the relevance score, rank the files in either ascendingly or descendingly. Different models have been proposed to score and rank files in IR community. Here we adopt the most widely used one tf-idf weighting consisting of two attributes-term frequency and inverse document frequency.

## Vector Space Model

While tf-idf depicts the weight of a single keyword on a file and the vector space model to score a file on multi keyword. The vector space model is an algebraic model for representing a file as a vector. The dimension of the vector corresponds to a separate term, the value in the vector is nonzero if a term occurs in the file, zero otherwise. The vector space model supports non binary presentation and multiterm. It allows computing a continuous degree of similarity between queries and files, according to their relevance they rank the files. It meets our needs of top-k retrieval. Files can be ranked in order the most relevant files can be found therefore.

### TRSE Design

Existing SSE schemes employ server-side ranking based on OPE to improve the efficiency encrypted cloud data retrieval. server-side ranking based on OPE violates the privacy of sensitive information, security cannot be tradeoff for efficiency. Ranking has to be left to the user side to achieve data privacy. Traditional user-side schemes due to the interaction between the server and the user on the user side, however, load heavy computational burden and high communication overhead including searchable index return and ranking score calculation. So the user-side ranking schemes are challenged by practical use. A better solution to privacy issues on the user side is by another server side scheme.

Novel technologies in cryptography community and IR community are employed on the user side, including homomorphic encryption and the vector space model. So in this scheme proposed, the data owner encrypts the searchable index with homomorphic encryption. A query consisting of multikeywords Received from the cloud server, it computes the scores from the encrypted index stored on cloud and then returns the encrypted scores of files to the data user. Data user decrypts the scores and picks out the top-k highest scoring files identifiers on the user side, to request to the cloud server. The cloud server and the data user takes a two-round communication between them for retrieval process. Thus the TRSE scheme, in which ranking is done at the user side while scoring calculation is done at the server side.

### 2.6. Secure Outsourcing ,Updation and Retrieval of Data over Cloud

The issues remain to be addressed in secure multi keyword top-k retrieval over encrypted cloud data. They are

### Efficiency Improvement

The main appeal of the modified FHEI that we employ in the TRSE scheme is its conceptual simplicity compared to Gentry's. At the cost of a large key size, this simplicity is achieved. Optimizations like modular reduction and compression can be employed to reduce the size of cipher text. For the practical system, the key size is still too large. The user encrypts his trapdoor and sends the cipher text to the cloud server. Therefore if the encrypted trapdoor size is too large, the communication overhead will be very high. To solve this

problem, a tradeoff of the security of search pattern may be needed unless a new encryption scheme that provides more reasonable cipher text size becomes available. From cryptography community, researchers have made several attempts to move toward practical fully homomorphic encryption over integers. Thus the efficiency of the TRSE scheme can be further improved.

### Enable Update

In a practical cloud computing system, a new challenge to the searchable encryption scheme is of data updates like adding or deleting files. Since data updates may be frequent, like for example every day in a medical system and users update their photo albums weekly or even daily and so it is necessary to consider the efficiency of update in searchable encryption design. Both the file itself and the searchable index require update operation in the presence of an update. In the TRSE scheme, The vector space model employed relies on the tf-idf weight. The idf factor may change for a keyword When a file is added or deleted. When updates occur, to avoid updating all the searchable index the file vectors should be independent to each other. The searchable index is built for each file, to store idf values for each keyword a possible solution is to only store tf values in file vectors and add another auxiliary vector. In this way, rather than all searchable index, update is limited to the auxiliary vector. The overall efficiency is almost immune to the update, Since the calculation is on the server side and the computing power on the server side is high.

### III. CONCLUSION

In this paper, we motivate and solve the problem of updation in secure multi keyword top-k retrieval over encrypted cloud data. We devise a server-side ranking SSE scheme, based on OPE invisibly leaking sensitive information. We then propose a TRSE scheme employing the fully homomorphic encryption that fulfills the security requirements of multi keyword top-k retrieval over the encrypted cloud data. We show that the proposed scheme guarantees data privacy based on OPE invisibly leaking sensitive information. It is necessary to allow multi keyword in the search request, considering the large number of data users and documents in the cloud, and return the most relevant documents in the order of their relevancy with these keywords. The two-round communication subdivides the Retrieval phase with two additional stages and thus introduces extra overhead. TRSE vastly reduces the communication overhead and the computing burden on the user side.

The overall efficiency is almost immune to the update Since the calculation is on the server side and the computing power on the server side is high. Researchers from cryptography community have made several attempts to move toward practical fully homomorphic encryption over integers. The efficiency of the TRSE scheme can be further improved by indicating this progress.

## REFERENCES

- [1] D. Boneh et al, Public- Key Encryption with Keyword Search, *Proc. Int'l Conf. "Theory and Applications of Cryptographic Techniques"* , vol. 53, no. 4, pp. 50-58, 2004.
- [2] C. Wang et al, Secure Ranked Keyword Search over Encrypted Cloud Data , *Proc. IEEE 30th Int'l Conf. Distributed Computing Systems (ICDCS)*, vol. 52, no. 4, pp. 748-764, 2010.
- [3] A. Swaminathan et al, Confidentiality-Preserving Rank-Ordered Search, *Proc. Workshop Storage Security and Survivability*, pp. 24-43, 2007.
- [4] P.olle et al, Secure Conjunctive Keyword Search over Encrypted Data, " *Proc. Second Int'l Conf. Applied Cryptography and Network Security* , pp. 31-45, 2004.
- [5] Jiadi Yu et al, " Toward Secure Multikeyword Top-k Retrieval over Encrypted Cloud Data," *IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING*, VOL. 10, NO. 4, JULY/AUGUST 2013
- [6] M. van Dijk et al, Fully Homomorphic Encryption over the Integers, *Proc. 29th Ann. Int'l Conf. Theory and Applications of Cryptographic Techniques*, pp. 24-43, 2010.
- [7] O. Regev et al, New Lattice-Based Cryptographic Constructions, *J. ACM*, vol. 51, no. 6, pp. 899-942, 2004.
- [8] P. Golle et al, Secure Conjunctive Keyword Search over Encrypted Data, " *Proc. Second Int'l Conf. Applied Cryptography and Network Security* ", pp. 31-45, 2004.