An Efficient Multi Keyword Ranked Search over Encrypted Cloud Data

Mrs. Bheema Rasagna & Mamillapalli Sadhana
#1 Assistant Professor,Dept of CSE, Visvodhya Engineering College,Kavali,India
#2Studnet,Dept of Master of Computer Application (MCA), PBR Visvodaya Institute Of Technology And Science ,Kavali,India

ABSTRACT-*Recently, more and more people are interested to outsource their local data to public cloud servers for great convenience and reduced costs in data management and security. But in fact of consideration privacy issues, sensitive data should be encrypted here before outsourcing, which obsoletes traditional data utilization like keyword-based document retrieval policy. In this project proposed to make the clusters of similar documents based on the cosine values of the document vectors. We also proposed a MRSE model used to search the documents which are in encrypted form. The proposed search technique only finds the cluster of documents with the highest similarity value instead of searching on the whole dataset. Processing the dataset on two algorithms shows that the time needed to form the clusters in the proposed method is less. When the documents in the dataset increases, the time needed to form clusters also increases. The result of the search shows that increasing the documents also increases the search time of the proposed method.*

KEYWORDS: Cloud computing, Encrypted data, Multi keyword search, Ranked Search, Similarity Matching,

1.INTRODUCTION

Cloud computing becomes popular as it provides huge storage space and high quality services. The large amount of data is created per day. It is a difficult task for the owner of the data to store and manage this large amount of data. To overcome this difficulty, the data owners can store their data on the cloud server to use the on demand applications and services from shared resources [1]. The cloud server providers agreed that their cloud service is armed with strong security constraints though security and privacy are major hindrances which avoid the use of cloud computing services [2]. To protect the sensitive data on the cloud server from unauthorized users, the data owners may encrypt the documents and uploads to cloud server [3]. In the earlier various strong cryptography methods were used to design the search techniques on the cipher text [4], [5],. These techniques needs many operations and require large amount of time. So these techniques are not suitable for big data where information volume is huge. The property of a document depends on its association The results of search returned to the users may contain damaged information due to hardware failure or storage corruption. Thus a mechanism should be given to users to check the accuracy of the search results.The proposed architecture of search technique is based on the cosine similarity clustering which maintain the association between plain text and encrypted text to improve the efficiency of search.

2. LITERATURE SURVEY

Chi Chen and Xiaojie Zhu used a hierarchical clustering method to maintain the close relationship between plain documents and encrypted documents to increase search efficiency within a big data environment. They also used a coordinate matching technique [5] to measure the relevance score between query and document. They did a model for the efficient multi-keyword ranked search and maintain the privacy of documents, rank security and relevance between retrieved documents.

Jiadi Yu and Peng Lu focused on the problems of the cipher text search using Searchable Symmetric Encryption (SSE) . This SSE technique helps data users to retrieve the documents over the encrypted documents. In Two Round Searchable Encryption (TRSE), they used the similarity relevance concept to solve the privacy issues in searchable encryption. They also showed server side ranking according to order preserving encryption (OPE).

N. Cao, C. Wang and M. Li used "inner product similarity" concept which can find the similarity measure of the information and the keywords of search.

Ruksana Akter, Yoojin Chung defined an evolutionary approach based on cosine similarity clustering. A document vector is used to create the index of every document. The cosine values between the document vectors are calculated. Clusters of the most relevant documents are formed on the basis of the cosine values. Another good feature of their work.

3. PROPOSED SCHEME

In this section, we give a detailed description of our scheme. We firstly propose to implement the semantic multi-keyword ranked search.

A. Our Scheme

As an effort towards the issue, in this paper, we propose an efficient multi-keyword ranked search scheme over encrypted mobile cloud data (MRSE) through blind storage. Our main contributions can be summarized as follows:

- We introduce a relevance score in searchable encryption to achieve multi-keyword ranked search over the encrypted mobile cloud data. In addition to that, we construct an efficient index to improve the search efficiency.

- By modifying the blind storage system in the MRSE, we solve the trapdoor unlinkability problem and conceal access pattern of the search user from the cloud server.
- We give thorough security analysis to demonstrate that the EMRS can reach a high security level including confidentiality of documents and index, trapdoor privacy, trapdoor unlinkability, and concealing access pattern of the search user. Moreover, we implement extensive experiments, which show that the EMRS can achieve enhanced efficiency in the terms of functionality and search efficiency compared with existing proposals.

B. Security Requirements

Specifically, the MRSE aims to provide the following four security requirements:

- **Confidentiality Of Documents And Index:** Documents and index should be encrypted before being outsourced to a cloud server. The cloud server should be prevented from prying into the outsourced documents and cannot deduce any associations between the documents and keywords using the index.

• **Trapdoor Privacy:** Since the search user would like to keep her searches from being exposed to the cloud server, the cloud server should be prevented from knowing the exact keywords contained in the trapdoor of the search user.

• **Trapdoor Unlinkability:** The trapdoors should not be linkable, which means the trapdoors should be totally different even if they contain the same keywords. In other words, the trapdoors should be randomized rather than determined. The cloud server cannot deduce any associations between two trapdoors.

• **Concealing Access Pattern Of The Search User:** Access pattern is the sequence of the searched results. In the EMRS, the access pattern should be totally concealed from the cloud server. Specifically, the cloud server cannot learn the total number of the documents stored on it or the size of the searched document even when the search user retrieves this document from the cloud server.

C. Blind Storage System

A blind storage system is built on the cloud server to sup- port adding, updating and deleting documents and concealing the access pattern of the search user from the cloud server. In the blind storage system, all documents are divided into fixed-size blocks. These blocks are indexed by a sequence of random integers generated by a document-related seed. In the view of a cloud server, it can only see the blocks of encrypted documents uploaded and downloaded. Thus, the blind storage system leaks little information to the cloud server. Specifically, the cloud server does not know which blocks are of the same document, even the total number of the documents and the size of each document. Moreover, all the documents and index can be stored in the blind storage system to achieve a searchable encryption scheme

4.PERFORMANCE EVALUATION

A. Functionality

Considering a large number of documents and search users in a cloud environment, searchable encryption schemes should allow privacy-preserving multi-keyword search and return documents in a order of higher relevance to the search request. As shown in TABLE 1, we compare functionalities among the EMRS, Cash‟s scheme, Cao‟s scheme and Naveed‟s scheme.

	[1]	[2]	[3]	MRSE
Multi-keyword	✓	✓		✓
Result Ranking		✓		✓
Relevance Scoring		✓		✓

B. Search efficiency

Search operation in Cao‟s scheme requires computing the relevance scores for all documents in the database. For each document, the cloud server needs to compute the inner product of two (d+2)-dimension vectors twice. Thus, the computation complexity for the whole data collection is O(md). As we can see, the search time in Cao‟s scheme linearly increases with the scale of the dataset, which is impractical for large-scale dataset. In the EMRS, by adopting the inverted index z which is built in the blind storage system, we achieve a sublinear computation overhead compared with Cao‟s scheme. Upon receiving stag, the cloud server can use stag to access blind storage and retrieve the encrypted relevance vector on the blocks indexed by the stag. These blocks consist of blocks of documents containing the stag-related keyword and some dummy blocks. Thus, the EMRS can significantly decrease the number of documents which are relevant to the searched keywords. Then, the cloud server only needs to compute the inner product of two (d+2)-dimension vectors for the associated documents rather than computing relevance scores for all documents as that in Cao‟s scheme . The computation complexity for search operation in the EMRS is O(α%sd), where %s represents the the number of documents which contain the keyword applied by the keyword-related token stag and the α is the extension parameter that scales the number of blocks in a document to the number of blocks in the set Sf . The value of %s can be small if the search user typically chooses the estimated least frequent keyword, such that the computation cost for search on the cloud server is significantly reduced. The computation cost of search phase is mainly affected by the number of documents in the dataset and the size of the keyword dictionary. In our experiments, we implement the index on the memory to avoid the time-cost I/O operations. Note that, although the time costs of search operation are linearly increasing in both schemes , the increase rate of the MRSE is less than half of that in Cao‟s scheme.

C. Measure

In this , we still use the measure of traditional information retrieval. Before the introduction of the F-measure‟s concept, we will firstly give the brief of the precision and recall. Precision is the fraction of retrieved instances that are relevant, while recall is the fraction of relevant instances that are retrieved. Both precision and recall are therefore based on an understanding and measure of relevance. F-measure that combines precision and recall is the harmonic mean of precision and recall. Here, we adopt F-measure to weigh the result of our experiments.

5.SECURITY ANALYSIS

A. **Confidentiality Of Documents And Index** The Documents Are encrypted by the traditional symmetric cryptography technique before being outsourced to the cloud server. Without a correct key, the search user and cloud server cannot decrypt the documents. As for index confidentiality, the relevance vector for each document is encrypted using the secret key M1, M2, and S. And the descriptors of the documents are encrypted using CP-ABE technique. And only the search user with correct attribute keys can decrypt the descriptor ABEυi(idi||Ki||x) to get the document id and the associated symmetric key. Thus, the confidentiality of documents and index can be well protected.

B. **Trapdoor Privacy** When a search user generates her trapdoor including the keyword-related token stag and encrypted query vector Q, she randomly chooses two numbers r and t.

Then, for the query vector q, the search user extends it as (rq,r,t) and encrypts the query vector using the secret key M1,M2 and S. Without the secret key M1,M2, S and K9, the cloud server cannot pry into the trapdoor. Thus, the keyword information In the trapdoor is totally concealed from the cloud server in the EMRS and trapdoor privacy is well protected.

C. **Trapdoor Unlinkability** Trapdoor unlinkability is defined as that the cloud server cannot deduce associations between any two trapdoors. Even though the cloud server cannot decrypt the trapdoors, any association between two trapdoors may lead to the leakage of the search user‟s privacy. We consider whether the two trapdoors including stag and the encrypted query vector Q can be linked to each other or to the keywords.

D. **Concealing Access Pattern Of The Search User** The access pattern means the sequence of the searched results .In Cash‟s scheme and Cao‟sscheme, the search user directly obtains the associated documents from the cloud server, which may reveal the association between the search request and the documents to the cloud server. In the EMRS by modifying the blind storage system, access pattern is well concealed from the cloud server.

TABLE 2- COMPARISON OF SECURITY LEVEL

	[1]	[2]	[3]	MRSE
CONFIDENTIALITY	✓	✓	✓	✓
TRAPDOOR UNLINKABILITY		✓		✓
CONCEAL ACCESS PATTERN OF SEARCH USER			✓	✓

6.CONCLUSION

In this paper, have proposed a multi-keyword ranked search scheme to enable accurate, efficient and secure search over encrypted mobile cloud data. Security analysis have demonstrated that proposed scheme can effectively achieve confidentiality of documents and index, trapdoor privacy, trapdoor unlinkability, and concealing access pattern of the search user. Extensive performance evaluations have shown that the proposed scheme can achieve better efficiency in terms of the functionality and computation overhead compared with existing ones. For the future work, will investigate on the authentication and access control issues in searchable encryption technique and provide an block insertion method to split the files and provide unique identification method and using decryption key download the files in the users side. This method can achieve the search efficiency.

REFERENCES

[1] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, „„Privacy-preserving multi- keyword ranked search over encrypted cloud data,"" IEEE Trans. Parallel Distrib. Syst., vol. 25, no. 1, pp. 222–233, Jan. 2014.

[2] W. K. Wong, D. W. Cheung, B. Kao, and N. Mamoulis, „„Secure kNN computation on encrypted databases,"" in Proc. ACM SIGMOD Int. Conf. Manage. Data, 2009, pp. 139–152.

[3] M. Naveed, M. Prabhakaran, and C. A. Gunter, „„Dynamic searchable encryption via blind storage,"" in Proc. IEEE Symp. Secur. Privacy, May 2014, pp. 639–654.

[4] H. Pang, J. Shen, and R. Krishnan, „„Privacy-preserving similarity-based text retrieval,"" ACM Trans. Internet Technol., vol. 10, no. 1, p. 4, 2010.

[5] D. Cash, S. Jarecki, C. Jutla, H. Krawczyk, M.-C. Roşu, and M. Steiner, „„Highly-scalable searchable symmetric encryption with support for Boolean queries,"" in Proc. CRYPTO, 2013, pp. 353–373.

Author's Profile

Mrs Bheema Rasagna received her M.Tech from JNTUA in 2012 . At Present She Is Working As Assistant Professor In The Department Of CSE. Visvodaya Engineering College, Kavali. She Has Total 10 Years Of Experience In Teaching. She Has Published 15 Papers In Reputed Journals And International Conference

Mamillapalli Sadhana pursuing Master of Computer Application (MCA) from PBR Visvodaya Institute of Technology and Science, Kavali, Nellore (dt),Andhra Pradesh