INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATIONS AND ROBOTICS
ISSN 2320-7345

PRIVACY PRESERVING MULTI KEYWORD RANKED SEARCH OVER ENCRYPTED DATA

[1] R.SHENBAGAVALLI, [2]M.KANIMOZHI, [3]P.GEETHA

[1]Associate Professor in Krishnasamy College of Engineering and Technology, Cuddalore, Tamilnadu.
E-mail id: shenbagasuresh@gmail.com
[2]M.E in Krishnasamy College of Engineering And Technology, Cuddalore, Tamilnadu.
E-mail id: kani071092@gmail.com
[3]M.E in Krishnasamy College of Engineering And Technology, Cuddalore, Tamilnadu.
E-mail id: geethaparthy@gmail.com

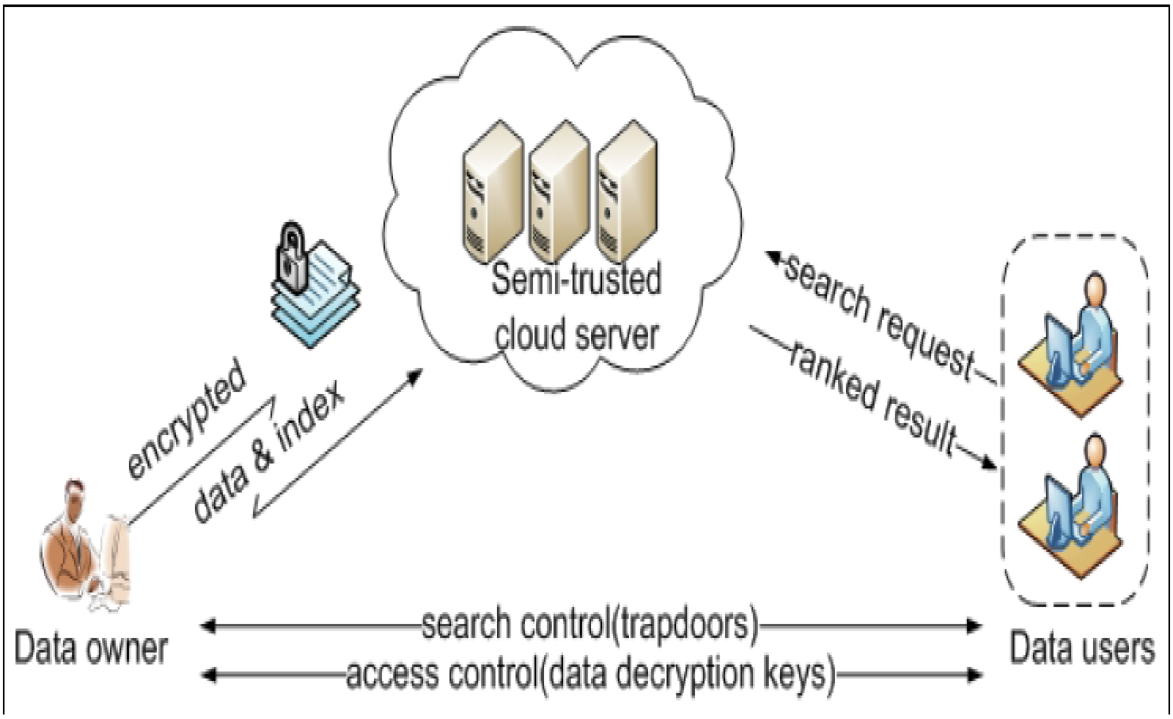
Abstract: - With the advent of cloud computing, data owners are motivated to outsource their complex data management systems from local sites to the commercial public cloud for great flexibility and economic savings. But for protecting data privacy, sensitive data have to be encrypted before outsourcing, which obsoletes traditional data utilization based on plaintext keyword search. Thus, enabling an encrypted cloud data search service is of paramount importance. In this project, for the first time, to define and solve the challenging problem of privacy-preserving Multi-keyword Ranked Search over Encrypted data (MRSE) in cloud computing, and to establish a set of strict privacy requirements for such a secure cloud data utilization system. Among various multi-keyword semantics, and to choose the efficient similarity measure of "coordinate matching," i.e., as many matches as possible, to capture the relevance of data documents to the search query. Further use "inner product similarity" to quantitatively evaluate such similarity measure. In this project first to propose a basic idea for the MRSE based on secure inner product computation, and then give two significantly improved MRSE schemes to achieve various stringent privacy requirements in two different threat models. To improve search experience of the data search service, further extend the two schemes to support more search semantics. 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, to investigate on the authentication and access control issues in searchable encryption technique.

Key terms: MRSE, Blind Storage, Index F

1. INTRODUCTION

Mobile cloud computing gets rid of the hardware limitation of mobile devices by exploring the scalable and virtualized cloud storage and computing resources, and accordingly is able to provide much more powerful and scalable mobile services to users. In mobile cloud computing, mobile users typically outsource their data to external cloud servers, e.g., iCloud, to enjoy a stable, low-cost and scalable way for data storage and access. However, as outsourced data typically contain sensitive privacy information, such as personal photos, emails, etc., which would lead to severe confidentiality and privacy violations, if without efficient protections. It is therefore necessary to encrypt the sensitive data before outsourcing them to the cloud. The data encryption, however, would result in salient difficulties when other users need to access interested data with search, due to the difficulties of search over encrypted data. This fundamental issue in mobile cloud computing accordingly motivates an extensive body of research in the recent years on the investigation of search- able encryption technique to achieve efficient searching over outsourced encrypted data .

ARCHITECTURE



EXISTING SYSTEM

In existing method data owner can be encrypted the documents and the cloud server, documents can be decrypted in cloud. the In order to meet the practical search requirements, search over encrypted data should support the following three functions. First, the searchable encryption schemes should support multi-keyword search, and provide the same user experience as searching in Google search with different keywords; single-keyword search is far from satisfactory by only returning very limited and inaccurate search results. Although the existing schemes aim at providing integrity verification for different data storage systems, the problem of supporting both public auditability and data dynamics has not been fully addressed. How to achieve a secure and efficient design to seamlessly integrate these two important components for data storage service remains an open challenging task in Cloud Computing.

DRAWBACKS

- Especially to support block insertion, which is missing in most existing schemes.
- Security will provided at the time of uploading only.

PROPOSED SYSTEM

We define and solve the challenging problem of privacy-preserving multi-keyword ranked search over encrypted cloud data (MRSE), and establish a set of strict privacy requirements for such a secure cloud data utilization system to become a reality. Among various multi-keyword semantics, we choose the efficient principle of “coordinate matching”. Here we are providing better security in owner’s upload side as well as on the download side. For better security client splitting that single file into nine different blocks and providing a unique identification number for each block.

2. 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.

3. 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.

4. 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			✓	✓

3. RELATED WORK

Searchable encryption is a promising technique that provides the search service over the encrypted cloud data. It can mainly be classified into two types: Searchable Public-key Encryp- tion (SPE) and Searchable Symmetric Encryption (SSE). Cash et al. adopt the inverted index TSet, which maps the keyword to the documents containing it, to achieve efficient multi-keyword search for large-scale datasets. Naveedet.al constructablindstoragesys- tem to achieve searchable encryption and conceal the access pattern of the search user. However, only single-keyword search is supported. Here this propose a security analysis with trapdoor privacy system and provide an security in download side.

4. 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

M.KANIMOZHI completed her B.E degree in Krishnasamy college of engineering and Technology, Cuddalore, Tamilnadu, India. Currently pursuing her M.E degree in computer science and Engineering from Krishnasamy college of engineering and Technology, Cuddalore, Tamilnadu, India. Her research interests are in the areas is cloud computing and image processing.

R. SHENBAGAVALLI completed her B.E degree in computer science in V.R.S college of engineering and technology, Villupuram, India. She had completed her M.E degree in computer science from SRM engineering college, Chennai, Tamilnadu. Her Area of interest is Networking, Cloud Computing and Analysis Algorithm.

P.GEETHA completed her B.E degree in computer science from V.R.S College of engineering and technology, Villupuram, India. Currently pursuing her M.E degree in computer science and Engineering from Krishnasamy college of engineering and Technology, Cuddalore, Tamilnadu, India. Her research interests are in the areas is cloud computing and image processing.