ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

A Survey on Developing Secure, Efficient and Dynamic Multi-Keyword Ranked Search Scheme over Encrypted Cloud Data

Kunti Matkar, Rekha Jadhav

Master of Computer Engg, G. H. Raisoni Collage of Engg and Technology, Wagholi, Pune, India[1]

HOD, Dept. of Information Technology, G. H .Raisoni Collage of Engg and Technology, Wagholi, Pune, India[2]

A**BSTRACT:** 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. Considering the large number of data users and documents in the cloud, it is necessary to allow multiple keywords in the search request and return documents in the order of their relevance to these keywords. Related works on searchable encryption focus on single keyword search or Boolean keyword search, and rarely sort the search results. In this paper, for the first time, we define and solve the challenging problem of privacy-preserving multi-keyword ranked search over encrypted data in cloud computing (MRSE). We establish a set of strict privacy requirements for such a secure cloud data utilization system. Among various multi-keyword semantics, we 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. We further use "inner product similarity" to quantitatively evaluate such similarity measure. We first 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 requirements in two different threat models. To improve search experience of the data search service, we further extend these two schemes to support more search semantics. Thorough analysis investigating privacy and efficiency guarantees of proposed schemes is given. Experiments on the real-world data set further show proposed schemes indeed introduce low overhead on computation and communication.

KEYWORDS: Searchable encryption, multi-keyword ranked search, dynamic update, cloud computing

I. INTRODUCTION

Computing resources are shared by many users. The benefits of cloud can be extended from individual users to organizations. The data storage in cloud is one of among them. The virtualization of hardware and software resources in cloud nullifies the financial investment for owning the data warehouse and its maintenance. Many cloud platforms like Google Drive, cloud, Sky Drive, Amazon S3, Drop box and Microsoft Azure provide storage services. Security and privacy concerns have been the major challenges in cloud computing. The hardware and software security mechanisms

Like firewalls etc. have been used by cloud provider. These solutions are not sufficient to protect data in cloud from unauthorized users because of low degree of transparency. Since the cloud user and the cloud provider are in the different trusted domain, the outsourced data may be exposed to the vulnerabilities. Thus, before storing the valuable data in cloud, the data needs to be encrypted.

To preserve the data privacy we need to design a searchable algorithm that works on encrypted data. Many researchers have been contributing to searching on encrypted data. The search techniques may be single keyword search or multi keyword search. In huge database the search may result in many documents to be matched with keywords. This causes difficulty for a cloud user to go through all documents and have most relevant documents. Search based on ranking is another solution, wherein the documents are ranked based on their relevancy to the keywords.

The researchers combined the rank of documents with multiple keyword searches to come up with efficient economically viable searchable encryption techniques. In searchable encryption related literature, computation time and

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

computation overhead are the two most frequently used parameters by the research in the domain for analysing the performance of their schemes.

This paper proposes a secure tree-based search scheme over the encrypted cloud data, which supports multi-keyword ranked search and dynamic operation on the document collection. Specifically, the vector space model and the widely-used "term frequency (TF) _ inverse document frequency (IDF)" model are combined in the index construction and query generation to provide multi-keyword ranked search. In order to obtain high search efficiency, we construct a tree-based index structure and propose a "Greedy Multi-keyword Boolean search allows the users to input multiple query keywords to request suitable documents. Our contributions are summarized as follows:

1) We design a searchable encryption scheme that supports both the accurate multi-keyword ranked search and flexible dynamic operation on document collection.2) Due to the special structure of our tree-based index, the search complexity of the proposed scheme is fundamentally kept to logarithmic. And in practice, the proposed scheme can achieve higher search efficiency by executing our "Greedy Depth-first Search" algorithm. Moreover, parallel search can be flexibly performed to further reduce the time cost of search process.

II. RELATED WORK

Ranked search can enable quick search of the most relevant data. Sending back only the top-k most relevant documents can effectively decrease network traffic. Some early works have realized the ranked search using order-preserving techniques, but they are designed only for single keyword search.

In literature many research works are not efficient in searching, especially for complex queries. This inefficiency may lead to leakage of valuable information to unauthorized peoples. Song et al, proposed symmetric searchable method based on cryptography which is practicable. The disadvantage of this method is that the word frequency will be known.

In order to overcome this, Goh et al proposed secure index table is constructed using pseudorandom functions which also uses unique document identifier randomized bloom filters. Bloom filters may lead to false positives.

Hence, Bosch et al used wild card searches for the concept
of Goh et al. To make the scheme more secured, Chang"s et al developed an index for each document, so that word count in a file is not disclosed. But, it is less efficient and do ot have the provision of random updates with new words.

The schemes Symmetric Searchable Encryption (SSE), Dynamic SSE (DSSE), are based on single keyword search w here addition and deletion of documents can be performed in index table.

Boneh et al proposed public key encryption with keyword search (PEKS) scheme has a drawback of an inference attack. The public key encryption methods are highly computational and complex that makes these algorithms inefficient. In Yang et al scheme the encrypted data is searched by individual users using a unique key allotted to them. The existing methods support only exact keyword search in the cloud.

Curtmola et al, proposed a method where an inverted index having document identifiers is maintained for each keyword.

The efficiency of the above scheme improvement is proposed by top-k single keyword retrieval schemes in the literature[17]. S.Buyrukbilen et al, proposed multi-keyword searches on public-key encrypted data.

bloom filter is used to encrypt keyword of each document and uses homomorphic encryption for the individual indexes into a tree structure. The query is hidden from the server or cloud provider by using an efficient private information retrieval (PIR) protocol.

MRSE scheme uses similarity based ranking which is created on the basis of term frequency and vector space. Search index is used for multi keyword search and ranking the search result. Synonym based multiple keywords ranked search over encrypted cloud data using balanced binary tree is proposed in. Here authors used symmetric encryption method for designing searchable encryption scheme and used b-tree for indexing. Many researchers have been investigating to identify a suitable privacy preserving technique for cloud domain, none of these solutions guarantee 100 percent privacy. There exists a wide range of research challenges. We therefore chose to work towards meeting this challenge.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

III. PROPOSED ALGORITHM

We first describe the unencrypted dynamic multi-keyword ranked search (UDMRS) scheme which is constructed on the basis of vector space model and KBB tree. Based on the UDMRS scheme, two secure search schemes (BDMRS and EDMRS schemes) are constructed against two threat models, respectively

1. Index Construction of UDMRS Scheme

We have briefly introduced the KBB index tree structure, which assists us in introducing the index construction. In the process of index construction, we first generate a tree node for each document in the collection. These nodes are the leaf nodes of the index tree. Then, the internal tree nodes are generated based on these leaf nodes. The formal construction process of the index is presented in

Algorithm 1

BuildIndexTree(F)

Input: the document collection F ={f1, f2, . . . , fn} with the identifiers FID = { FID|FID =1, 2, . . . , n}

Output: the index tree T
1: for each document fFID in F do
2: Construct a leaf node u for fFID, with u:ID =GenID(), u:Pl = u:Pr = null, u:FID = FID, and D[i]=TFfFID;wi for i = 1, . . .,m;
3: Insert u to CurrentNodeSet;
4: end for
5: while the number of nodes in CurrentNodeSet is larger than 1 do
6: if the number of nodes in CurrentNodeSet is even, i.e., 2h Then
7: for each pair of nodes u' and u'' in CurrentNodeSet do
8: Generate a parent node u for u0 and u00, with u:ID = GenID(), u:Pl ¼=u', u:Pr = u'', u:FID = 0 and D(i)= max{u':D(i) , u'' :D(i) for each i = 1, . . .,m;
9: Insert u to TempNodeSet;
10: end for
11: else
12: for each pair of nodes u0 and u00 of the former (2h _ 2) nodes in CurrentNodeSet do
13: Generate a parent node u for u0 and u00;
14: Insert u to TempNodeSet;
15: end for

16: Create a parent node u1 for the (2h-1) th and 2hth node, and then create a parent node u for u1 and the (2h + 1)th node;
17: Insert u to TempNodeSet;
18: end if
19: Replace CurrentNodeSet with TempNodeSet and then clear TempNodeSet;
20: end while

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

21: return the only node left in CurrentNodeSet, namely, the root of index tree T ;

2. Search Process of UDMRS Scheme

The search process of the UDMRS scheme is a recursive procedure upon the tree, named as "Greedy Depth-first Search" algorithm. We construct a result list denoted as RList, whose element is defined as RScore; FIDi. Here, the RScore is the relevance score of the document fFID to the query, which is calculated according to Formula (1). The RList stores the k accessed documents with the largest relevance scores to the query. The elements of the list are ranked in descending order according to the RScore, and will be updated timely during the search process. Following are some other notations, and the GDFS algorithm is described in Algorithm 2.

Algorithm 2

GDFS (IndexTreeNode u)
1: if the node u is not a leaf node then
2: if RScore(Du,Q) > kthscore then
3: GDFS(u:hchild);
4: GDFS(u:lchild);
5: else
6: return
7: end if
8: else
9: if RScore(Du,Q) > kth score then
10: Delete the element with the smallest relevance score from RList;
11: Insert a new element (RScore(Du,Q); u:FID) and sort all the elements of RList;
12: end if
13: return
14: end if

IV. SIMULATION RESULTS

This paper proposes a tree-based search scheme over the encrypted cloud, which supports multi-keyword ranked search and dynamic operation on the document collection. Specifically, the vector space model and "term frequency (TF) × inverse document frequency (IDF)" model are collectively used in the index development and query generation to provide multikeyword ranked search. In order to obtain high search efficiency, a tree-based index structure is constructed and a "Greedy Depth-first Search" algorithm based on this index tree is proposed. Due to the special structure of our tree-based index, the proposed search scheme can achieve sub-linear search time flexibly and deal with the deletion and insertion of documents. The secure kNN algorithm is utilized to encrypt the index and query vectors, and meanwhile ensure accurate relevance score calculation between encrypted index and query vecto1rs. To resist different attacks in different threat models, we construct two secure search schemes: the basic dynamic multi-keyword ranked search (BDMRS) scheme in the known cipher text model, and the enhanced dynamic multi-keyword ranked search (EDMRS) scheme in the known background model.
The system model in this paper incorporates three unmistakable substances: data owner, data user and cloud server, as illustrated in Fig.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

Data owner
has a gathering of records F = {f1; f2; :::; fn} that he needs to outsource to the cloud server in encoded structure while up 'til now keeping the ability to check on them for convincing utilization. Data owner firstly manufactures a secure searchable tree index I from archive accumulation F, and a short time later makes an encrypted document gathering C for F. A brief span later, the data owner outsources the encoded accumulation C and the secure index I to the cloud server, and safely disseminates the key data of trapdoor era and document decryption to the approved data users. Additionally, the data owner observe his documents those are stored on cloud server. When updating, the data owner creates the upgradable data locally and sends it to the server.

Data users are approved ones to get to the archives of data owner. With t query keywords, the approved user can create a trapdoor TD as indicated by search control mechanisms to get k encrypted documents from cloud server. By then, documents are decrypt using shared secret key.

Cloud server stores the encrypted document accumulation C and the encrypted searchable tree index I for data owner. In the wake of tolerating the trapdoor TD from the data user, look over the index tree I, in conclusion gives back the relating gathering of top-k situated encoded reports. Also, in the wake of tolerating the update information from the data owner, the server needs to update the index I and document gathering C as per the received information. A Dynamic Key generation scheme is involved for achieving parallelism for the existing multi keyword ranked search scheme. Several steps are involved during sending data to the multiple users. Firstly, when there is a request for data by multiple users, each user will be issued a dynamic key for accessing the encrypted cloud data. Secondly, the user can access the data only when the appropriate dynamic key is entered. Hence this improves the privacy as well as achieves parallelism.

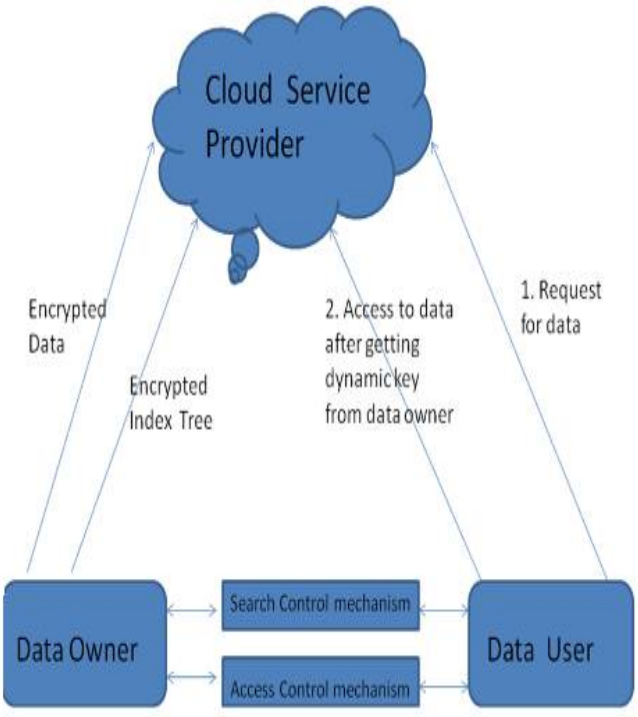

Fig. 1:- System Model

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

V. CONCLUSION AND FUTURE WORK

To design a dynamic Searchable encryption scheme whose updating operation can be completed by cloud server only while reserving the ability to support multi-keyword ranked search. To rebuild the index and distribute the new secure keys to all the authorized users. To improve the SE scheme of dishonest data user will lead to many secure problems.

REFERENCES

1. Xia, Zhihua, et al. "A secure and dynamic multi -keyword ranked search scheme over encrypted cloud data." IEEE Transactions on Parallel and Distributed Systems, vol. 27, no .2 , pp. 340-351,2016
2. C. Wang, N. Cao, K. Ren, and W. Lou, "Enabling secure and efficient ranked keyword search over outsourced cloud data," Parallel and Distributed Systems, IEEE Transactions on, vol. 23, no. 8, pp. 1467–1479, 2012.
3. N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-preserving multi-keyword ranked search over encrypted cloud data," in IEEE INFOCOM, April 2011, pp. 829–837.
4.O. Goldreich and R. Ostrovsky, "Software protection and simulation on oblivious rams," Journal of the ACM (JACM), vol. 43, no.3, pp. 431–473, 1996.
5. D. Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, „Public key encryption with keyword search," in Proc. of EUROCRYPT, 2004.
6. C. Wang et al., „Secure Ranked Keyword Search Over Encrypted Cloud Data," Proc. I CDCS "10, 2010
7. Wenjun Lu; Varna, A.L.; Min Wu, „Confidentiality- Preserving Image Search: A Comparative Study Between Homomorphic Encryption and Distance-Preserving Randomization," Access, IEEE, vol.2, no.3, pp.125-141, 2014
8. W. K. Wong, D. W. Cheung, B. Kao, and N. Mamoulis, „Secure knn computation on encrypted databases," in Proc. of SIGMOD, 2009
9. Wenjun Lu; Varna, A.L.; Min Wu, „Confidentiality- Preserving Image Search: A Comparative Study Between Homomorphic Encryption and Distance-Preserving Randomization," Access, IEEE, vol.2, no.3, pp.125-141, 2014
10. Kamara and K. Lauter, "Cryptographic cloud storage," in Financial Cryptography and Data Security Springer, pp. 136–149. 2010
11. C. Gentry, "A fully homomorphic encryption scheme," Ph.D. dissertation, Stanford University, 2009.