Scientific Journal of Impact Factor (SJIF): 4.72

e-ISSN : 2348-4470
p-ISSN : 2348-6406

International Journal of Advance Engineering and Research Development

Special Issue on Recent Trends in Data Engineering

Volume 4, Special Issue 5, Dec.-2017

Secure and Efficient Multi-keyword Ranked Search Method on Cloud

Prof. M.A.Zope[1], Chetna Raut[2], Nayana Dashmukhe[3], Neha Fegade[4], Anushree Gawande[5]

[1]*Department of Computer Engineering, AISSMS IOIT, Pune*
[2]*Department of Computer Engineering, AISSMS IOIT, Pune*
[3]*Department of Computer Engineering, AISSMS IOIT, Pune*
[4]*Department of Computer Engineering, AISSMS IOIT, Pune*
[5]*Department of Computer Engineering, AISSMS IOIT, Pune*

Abstract –*The extensive use of cloud services has resulted in dramatic growth in volume of data which has made information retrieval much more difficult than before. Even text documents are encrypted before being outsourced to cloud servers. This helps to protect user's data privacy. Existing techniques to search over encrypted data are not suitable for a huge data environment. Due to the blind encryption, relationship between the documents is concealed which further leads to search accuracy performance degradation. Therefore it is necessary to adopt an approach to support more search semantic and for fast search within enormous data. A hierarchical clustering method for cipher text search is proposed in this paper. The proposed approach clusters the documents based on the minimum similarity threshold, and then partitions the resultant clusters into sub-clusters until the constraint on the maximum size of cluster is reached. The very first thing which cloud server does is that it performs search activity and then selects the k documents (previously decided by the user and sent to the cloud server) from the minimum desired sub-category. To request desired documents, instead of using single keyword search query, multi-keyword search technique is proposed.*

Keywords- *Security, Data Encryption, Distributed system, Information search and Retrieval, Clustering*

I. INTRODUCTION

Present world scenario is that a huge amount of data is generated every day by different organizations. Generally these enterprises or organizations store their confidential data on cloud. The cloud service providers profess the safety of the stored data. But, Security and privacy of data still is compromised.

Basically to protect the data, it is first encrypted and then outsource to the cloud. But, retrieval of these data requires large amount of time which increases the overhead. It is nearly impossible or impractical in a way to retrieve all the encrypted documents by decrypting and downloading it locally. The traditional approach to search and retrieve data on plain text cannot be applied here.

In this proposed system, every document is represented as a point in a HD space. In other words, if the points whose distances are short in the HD space then they can be classified into a specific category. By comparing the documents which belong to dataset we can conclude that the targeted documents are less in number.

Thus, we can reduce searching time. Backtracking approach is more efficient than the traditional sequential search method, so it is used for searching the required documents. At the beginning, the cloud server will analyze the categories and accordingly select the category which is relevant. Based on the relevance the cloud server will choose top k documents from the selected sub-category. User defines the value of k and dispatches to cloud server. If current sub-category fails to satisfy the k documents, cloud server will back track to its parent and select the desired documents from its siblings. If selected sub-category is unable to meet the condition for k documents then it will directly trace back to its parent and process continues for the siblings as well. The terminating conditions for this are if the criteria (k) is satisfied or the root reached.

II. Traditional Approach

In traditional way we use data encryption to avoid information leakage. However, it is a very challenging task to search on encrypted data as server-side data utilization is more. In order to address the above problem, a general solution with fully-homomorphic encryption has been designed. Searchable encryption schemes allow users to store data in encrypted form and perform search over cypher text domain. Till now numerous works have been proposed under various threat models to perform different search functionality. Among them, multi keyword ranked search has more practical applicability.

Earlier approaches are not able to perform effective search on huge amount data. In addition, due to blind encryption the relationship between documents is concealed. To completely describe a document, the relationship between documents is important. Therefore, proposing a method which will be able to sustain and use this relationship to make searching fast is required.

III. SURVEY DETAILS

1. **An efficient privacy-preserving ranked keyword search method**
Chi Chen, et al

In this paper [1], they have examined searching on cipher text in storage of cloud. The proposed system have analyzed the problem of sustaining well-formed relationships among various plain text documents over related encrypted document give the design technique to increase the performance of semantic search.

2. **A Secure and Dynamic Multi-keyword Ranked Search Scheme over Encrypted Cloud Data**
Zhihua Xia,Xinhui Wang, Xing Ming Sun, Qian Wang

In this paper [2], the system for searching the most relevant documents over encrypted data relevance score is calculated. For index construction and query reformation combination of vector space model and TF-IDF model is proposed. To make searching efficient the given system has used Greedy Depth First Search (GDFS).

3**. Privacy-preserving similarity based text retrieval**
C. Wang, N. Cao, K. Ren, and W. J. Lou,

In this paper [3], the system tries to solve the problem of searching on the encrypted data stored. Even if perform searching on encrypted documents the chances of getting query related documents are very less. This paper partially overcomes it. It also assures the file retrieval accuracy. Hence without compromising the privacy the service is properly utilized.

4. **Privacy-preserving similarity based text retrieval**
H. Pang, J. Shen, and R. Krishnan

In this paper [4], the proposed work tries to hide search results from unauthorized observers. Also server side ranking provides query related results which reduces overhead of the end users.

5. **Privacy-preserving multi-keyword ranked search over encrypted cloud data**
N. Cao, C.Wang, M. Li, K. Ren, and W. J. Lou, "

In this paper[5], the challenging issue of preserving the privacy of multi-keyword ranked search over encrypted information in cloud computing (MRSE) is proposed. A set of strict privacy requirements are designed for a secure cloud information usage framework. Coordinate matching is used to capture the relevance of data documents to the search query. Even the inner product similarity is used to quantitatively assess such similarity measure. A fundamental idea for the MRSE based on secure inner product computation is proposed

6. **"Privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking**
W.Sun, B.Wang, N. Cao, M. Li, W. Lou, Y. T. Hou, and H. Li

In this paper[6] ,to support multi-keyword search and search result ranking, they propose to create the search index based on term frequency(TF) and also the vector area model with cosine similarity .This is performed to achieve the higher search result accuracy. To enhance the search efficiency a tree-based index structure is proposed . Furthermore they propose two secure index schemes to satisfy the privacy needs under strong threat models which include known cipher text model and known background model.

7. Secure multi-keyword search supporting dynamic update and ranked retrieval
Jingbo Yan; Yuqing Zhang; Xuefeng Liu
In the paper [7], to enhance the security function-hiding inner product encryption is proposed which prevents the leakage of search pattern. A tree-based index structure is adopted to facilitate the searching process and updating operations.An extensive security analysis is provided and experiments over the real world data show that the proposed scheme is efficient

IV. PROPOSED SYSTEM

A. Architecture:

Data owner:
The data owner performs following tasks-
1 selects the collection documents $d_i = \{d_1, d_2 \ldots d_m\}$ to be stored on cloud.
2. Builds the secure searchable index tree (I) of these documents.
3. Generate the encrypted document collection (C) for these documents.
4. At last uploads the index tree (i) and collection of encrypted documents(C) on cloud server.
5. Distributes the secret key (sk) to authorized user which is required to generate the trap door.
6. Also shares the decryption key (K) with the user.

Following things need to be considered while performing above taks are-
1. Maintain the capability to search on encrypted documents.
2. In case the documents are updated the required information is stored locally.

Data users
The authorized users can access the documents stored on cloud by data owner. They are the Data users.

With query which can be single keyword or multi-keywords (w), the authorized user can generate a trapdoor $T_{w.}$
This is done according to search control mechanisms to fetch "*k* "encrypted documents from cloud server. The trapdoor is forwarded to cloud. The value of k as mentioned earlier is user defined. After fetching the result ,the uer now decrypts the documents using the Key(K) provided by data owner

Cloud server
Cloud server stores the -
1. Encrypted document collection (C*)*
2. Encrypted searchable index tree *I* form data owner.
Upon receiving the trapdoor T_w from the data user
1. The cloud server executes search over the index tree *I*
2. Responds the corresponding collection of top- *k* ranked encrypted documents to the user.

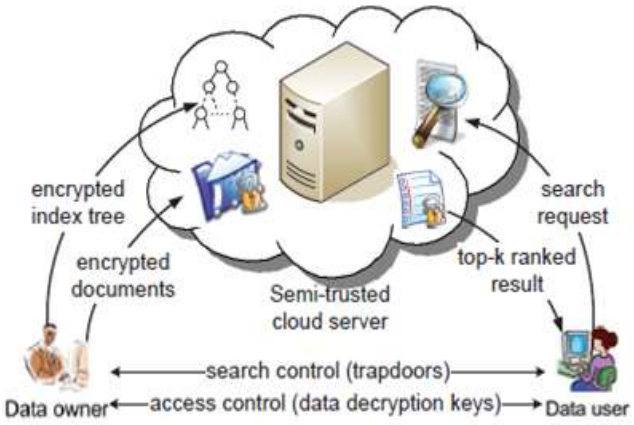

Fig.1- Ranked search over encrypted cloud data architecture

B. System Model:
MRSE-HCI architecture:
In Multi keyword Ranked Search Encryption –Hierarchical Clustering Index, each document can be represented using a vector with index. Keywords are represented as dimensions of the vector and their values represent whether the keyword is present in the document or not. Likewise, the search keywords are also represented in the same manner. Search results are generated based on the relevance scores which are figured out by finding the product the transformed query and documents. According to the relevance score the related top k documents are returned to the end user. MRSE-HCI architecture is shown in Fig. 2.

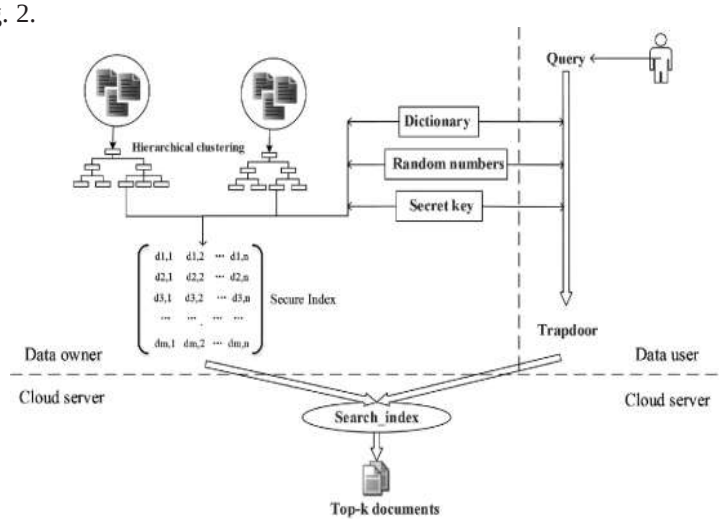

Fig.2- MRSE-HCI architecture

The data user submits a query to the cloud server for getting desired documents and the cloud server returns the target documents to the data user. This architecture mainly consists of following algorithms-

1) Key-generation:
   Two keys are generated-
   1. Secret key (sk) that is used to encrypt the index tree.
   2. key (K) used to encrypt the documents.
2) Index tree generation:
   The above mentions key "sk" is used to generate the encrypted index (I).
   Clustering process is also performed.

3) Encryption:
   This phase includes the encryption of the documents collection (D) by the "K".
4) Trapdoor:
   The query vector (Tw) is generated in this phase. This "Tw" is generated using keywords entered by users and the secret key (sk).
5) Search: In this phase, cloud server compares trapdoors (T_w) with index .Relevance score is calculated and based on it the top-k retrieval are retrieved.
6) Decryption: The retrieved documents are selected and decrypted accordingly using key (K) generated earlier.

V. EXPECTED RESULTS

Documents are clustered, encrypted and stored on Cloud. User requests the cloud by entering single or multi-keyword search query. The cloud server will respond to user's query with top K relevant documents as k is provided by the user. User will further decrypt required documents from the list provided by cloud server.

VI. CONCLUSION

This paper proposes a safe and efficient search scheme. It guarantees the accurate multi-keyword ranked search. We explored the problem of sustaining the relations among various plain documents over the correlated encrypted documents and give the design technique to improve the performance of the search on these documents. For data explosion and for retrieval of online information MRSE-HCI architecture is used. The documents retrieved by the user is also verified which maintains the integrity.

REFERENCES

[1] Chen, Chi, et al. "An efficient privacy-preserving ranked keyword search method." IEEE Transactions on Parallel and Distributed Systems 27.4 (2016): 951-963.

[2]Zhihua Xia, Xinhui Wang, Xingming Sunand Qian Wang "A Secure and Dynamic Multi-keyword Ranked Search Scheme over Encrypted Cloud Data" IEEE transactions on parallel and distributed systems vol: pp no: 99 year 2015

[3] C. Wang, N. Cao, K. Ren, and W. J. Lou, "Enabling secure and efficient ranked keyword search over outsourced cloud data," IEEE Trans. Parallel Distrib. System, vol. 23, no. 8, pp. 1467-1479, Aug. 2012.

[4] H. Pang, J. Shen, and R. Krishnan, "Privacy-preserving similarity based text retrieval," ACM Trans. Internet Technol., vol. 10, no. 1, pp. 39, Feb. 2010.

[5] N. Cao, C.Wang, M. Li, K. Ren, and W. J. Lou, "Privacy-preserving multi-keyword ranked search over encrypted cloud data," in Proc. IEEE INFOCOM, Shanghai, China, 2011, pp. 829-837.

[6] W. Sun, B.Wang, N. Cao, M. Li, W. Lou, Y. T. Hou, and H. Li, "Privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking," in Proc. 8th ACM SIGSAC Symp. Inform. Comput.Commun. Security, Hangzhou, China, 2013, pp. 71-82.

[7]Jingbo Yan; Yuqing Zhang; Xuefeng Liu,"Secure multi-keyword search supporting dynamic update and ranked retrieval", IEEE Trans. Parallel Distrib. System, vol. 13, no. 8, pp. 1467-1479, Nov. 2016.

[8] C. Chen, X. J. Zhu, P. S. Shen, and J. K. Hu, "A hierarchical clustering method For big data oriented cipher text search," in Proc. IEEE INFOCOM, Workshop on Security and Privacy in Big Data, Toronto, Canada, 2014, pp. 559-564.

[9] Cash, David, et al. "Dynamic Searchable Encryption in Very-Large Databases: Data Structures and Implementation." IACR Cryptology ePrint Archive 2014 (2014): 853.

[10] Yanzhu Liu, Zhi Li, Wang Guo and Wu Chaoxia, "Privacy-preserving multi-keyword ranked search over encrypted big data," Third International Conference on Cyberspace Technology (CCT 2015), Beijing, 2015, pp.1-3.

[11] Ching-Yang Tseng, Chang Chun Lu and Cheng-Fu Chou, "Efficient privacy preserving multi-keyword ranked search utilizing document replication and partition, 2015" 12th Annual IEEE Consumer Communications and Networking Conference (CCNC), Las Vegas, NV, 2015, pp. 671-676.

