OPEN ACCESS INTERNATIONAL JOURNAL OF SCIENCE & ENGINEERING

SECURE AND EFFICIENT RANKED KEYWORD COVER SEARCH IN CLOUD COMPUTING

Sandhyarani Galphade[1], Prof. D. R. Patil[2]

PG Student, Dept. of Computer Engineering, JSCOE, Hadapsar, Pune.[1]
Assistant Professor, Dept. of Computer Engineering, JSCOE, Hadapsar, Pune, India[2]

Abstract: ***An increasing popularity of cloud computing, an ever increasing number of information owners are inspired to outsource their information to cloud servers for awesome comfort and reduced cost in information administration. However, sensitive data should be encrypted before outsourcing for privacy requirements, which obsoletes data utilization like keyword-based document retrieval. In this paper, we introduce a secure multi-keyword ranked search scheme over encoded cloud information, which at the same time supports dynamic update operations like cancellation and addition of archives. In particular, the vector space demonstrate and the generally utilized TF-IDF show are joined in the record development and inquiry age. In this paper build a unique tree-based file structure and propose an "Greedy Depth-first Search" algorithm to give productive multi-keyword ranked search. The protected kNN algorithm is used to encode the file and inquiry vectors, what's more, in the interim guarantee precise importance score count between encrypted index and query vectors. Keeping in mind the end goal to oppose attacks, ghost terms are added to the index vector for blinding indexed lists. Because of the utilization of our unique tree-based index structure, the proposed plan can accomplish sub-direct inquiry time and manage the cancellation and addition of records adaptable. Extensive experiments are led to show the effectiveness of the proposed scheme.***

KEYWORDS: *TF-IDF, multi-keyword, KNN*

---∴ ∴ ∴---

I INTRODUCTION

CLOUD computing has been considered as a new model of enterprise IT infrastructure, which can organize huge resource of computing, storage and applications, and enable users to enjoy ubiquitous, convenient and on demand network access to a shared pool of configurable computing resources with great efficiency and minimal economic overhead. Attracted by these appealing features, both individuals and enterprises are motivated to outsource their data to the cloud, instead of purchasing software and hardware to manage the data themselves. Despite of the various advantages of cloud services, outsourcing sensitive information (such as e-mails, personal health records, company finance data, government documents, etc.) to remote servers brings privacy concerns. The cloud service providers (CSPs) that keep the data for users may access users' sensitive information without authorization. A general approach to protect the data confidentiality is to encrypt the data before outsourcing. However, this will cause a huge cost in terms of data usability. For example, the existing techniques on keyword-based information retrieval, which are widely used on the plain text data, cannot be directly applied on the encrypted data. Downloading all the data from the cloud and decrypt locally is obviously impractical. In order to address the above problem, researchers have designed some general purpose solutions with fully-homomorphism encryption or oblivious RAMs. However, these methods are not practical due to their high computational overhead for both the cloud sever and user. On the contrary, more practical special-purpose solutions, such as searchable encryption (SE) schemes have made specific contributions in terms of efficiency, functionality and security. Searchable encryption schemes enable the client to store the encrypted data to the cloud and execute keyword search over cipher text domain. So far, abundant works have been proposed under different threat models to achieve various search functionality, such as single keyword search, similarity search, multi-keyword Boolean search, ranked search, multi keyword ranked search, etc. Among them, multi-keyword ranked search achieves more and more attention for its practical applicability. Recently, some dynamic schemes have been proposed to support inserting and

deleting operations on document collection. These are significant works as it is highly possible that the data owners need to update their data on the cloud server. But few of the dynamic schemes support efficient multi-keyword ranked search.

II LITERATURE SURVEY

In this paper, we propose a semantic multi-keyword ranked search scheme over the encrypted cloud data, which simultaneously meets a set of strict privacy requirements. Firstly, we utilize the ”Latent Semantic Analysis” to reveal relationship between terms and documents. The latent semantic analysis takes advantage of implicit higher-order structure in the association of terms with documents (”semantic structure”) and adopts a reduced - dimension vector space to represent words and documents. Thus, the relationship between terms is automatically captured. Secondly, our scheme employ secure ”k-nearest neighbor(k-NN)” to achieve secure search functionality[1].

In this paper, we define and solve the problem of secure ranked keyword search over encrypted cloud data. Ranked search greatly enhances system usability by enabling search result relevance ranking instead of sending undi erentiated results, and further ensures the file retrieval accuracy. Specically, we explore the statistical measure approach, i.e., relevance score, from information retrieval to build a secure searchable index, and develop a one-to-many orderpreserving mapping technique to properly protect those sensitive score information. The resulting design is able to facilitate efficient server-side ranking without losing keyword privacy[2].

In this paper, for the first time we define and solve the problem of effiective yet secure ranked keyword search over encrypted cloud data. Ranked search greatly enhances system usability by returning the matching files in a ranked order regarding to certain relevance criteria (e.g., keyword frequency), thus making one step closer towards practical deployment of privacy-preserving data hosting services in Cloud Computing. We first give a straightforward yet ideal construction of ranked keyword search under the state-of-the-art searchable symmetric encryption (SSE) security denition, and demonstrate its in efficiency[3].

key-generation method for the underlying somewhat homomorphic encryption, that does not require full polynomial inversion. This reduces the asymptotic complexity from (n2.5) to (n1.5) when working with dimension-n lattices (and practically reducing the time from many hours/days to a few seconds/minutes). Other optimizations include a batching technique for encryption, a careful analysis of the degree of the decryption polynomial, and some space/- time trade-os for the fully-homomorphism scheme [4].

In this paper, we propose a practical privacy-preserving ranked key- word search scheme based on PIR that allows multi-keyword queries with ranking capability. The proposed scheme increases the security of the keyword search scheme while still satisfying efficient computation and communication requirements. To the best of our knowledge the majority of previous works are not efficient for assumed scenario where documents are large files[5].

III PROBLEM STATEMENT

To search the encrypted index and data files over the cloud. To improve the network bandwidth efficacy and usability of cloud data.

IV SYSTEM ARCHITECTURE

The system model involves three different entities: data owner, data user and cloud server, as illustrated in system architecture. Data owner has a collection of documents F ={ f1, f2 . . . fn} that he wants to outsource to the cloud server in encrypted form while still keeping the capability to search on them for effective utilization. In our scheme, the data owner first builds a secure searchable tree index I from document collection F, and then generates an encrypted document collection C for F. Afterwards, the data owner outsources the encrypted collection C and the secure index I to the cloud server, and securely distributes the key information of trapdoor generation (including keyword IDF values) and document decryption to the authorized data users. Besides, the data owner is responsible for the update operation of his documents stored in the cloud server. While updating, the data owner generates the update information locally and sends it to the server. Data users are authorized ones to access the documents of data owner. With t query keywords, the authorized user can generate a trapdoor TD according to search control mechanisms to fetch k encrypted documents from cloud server. Then, the data user can decrypt the documents with the shared secret key. Cloud server stores the encrypted document collection C and the encrypted searchable tree index I for data owner. Upon receiving the trapdoor TD from the data user, the cloud server executes search over the index tree I, and finally returns the corresponding collection of top-k ranked encrypted documents. Besides, upon receiving the update information from the data owner, the server needs to update the index I and document collection C according to the received

information. The cloud server in the proposed scheme is considered as “honest-but curious”,

which is employed by lots of works on secure cloud data search. Specifically, the cloud server honestly and correctly executes instructions in the designated protocol. Meanwhile, it is curious to infer and analyze received data, which helps it acquire additional information. Depending on what information the cloud server knows, we adopt the two threat

models proposed by Cao et al. Known cipher text model. In this model, the cloud server only knows the encrypted document collection C, the searchable index tree I, and the search trapdoor TD submitted by the authorized user. That is to say, the cloud server can conduct cipher text only attack (COA) in this model.

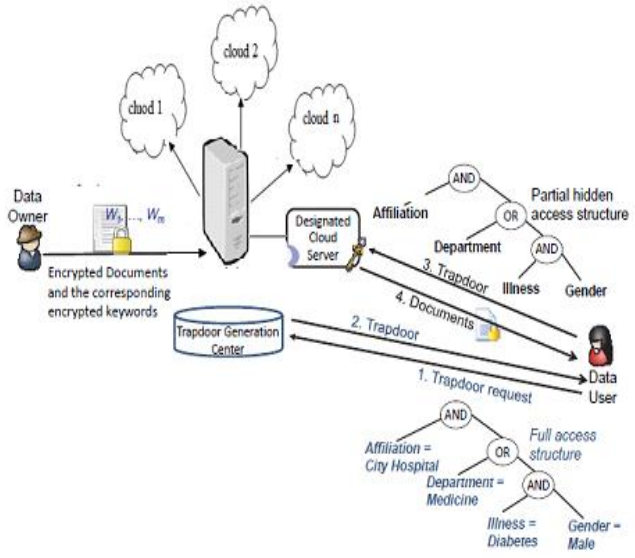

Figure 1: System architecture

V. ALGORITHM

Algorithm Ranked Serial Binary Search (RSBS) algorithm

Input: Noised trapdoors (one per search keyword): T1,.., T e

Encrypted document indexes: A = I1—-IN

The number of documents to return: k

Output:

Top-k documents that best match the search request: D = {D1.D2..,Dk}

1: Scores = zeros(0;N) // create an array of N zeros

2: for i := 1 to N do

3: for n := 1 to e do

4: Score[i] Score[i] + bsearch(Tn, Ii, 1 ,Si)// search if the keyword appears in any of the s slices of the document

5: end for

6: end for

7: sorted, indices = sort(Scores) // sort the score array and get the indices or old element in the sorted array.

8: D ¡– indice [0 : k - 1] // get the top-k documents

9: return D

VI. RESULT

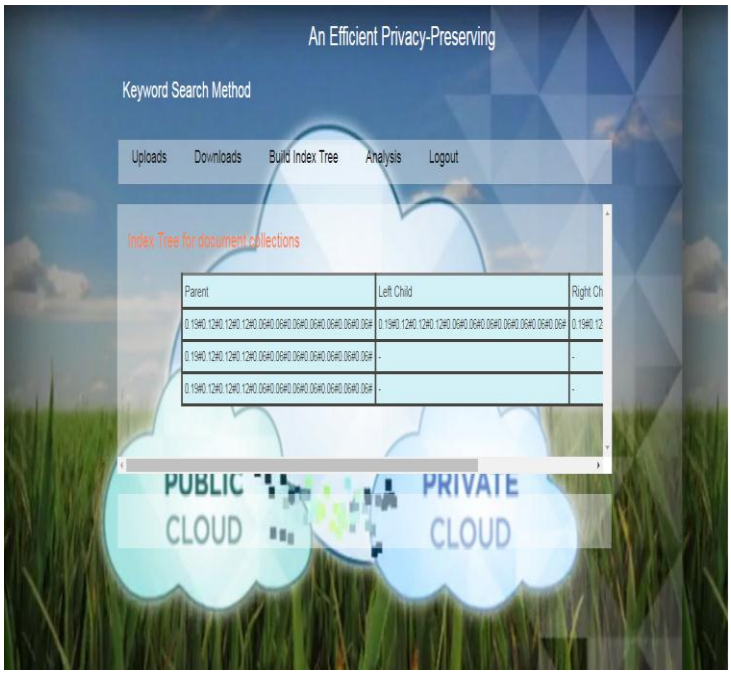

Figure 2: Index Tree For Document Collection

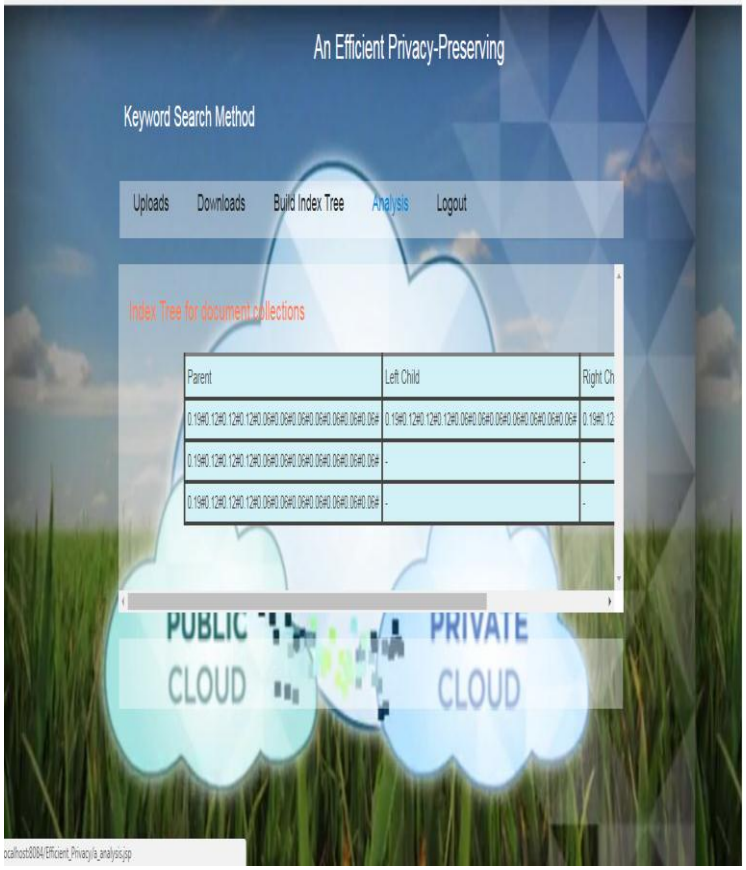

Figure 3: Analysis

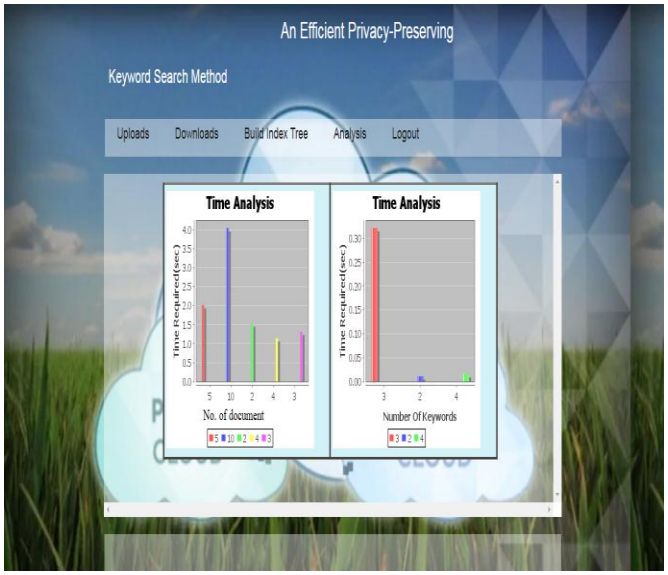

Figure 4: Time Analysis Ratio

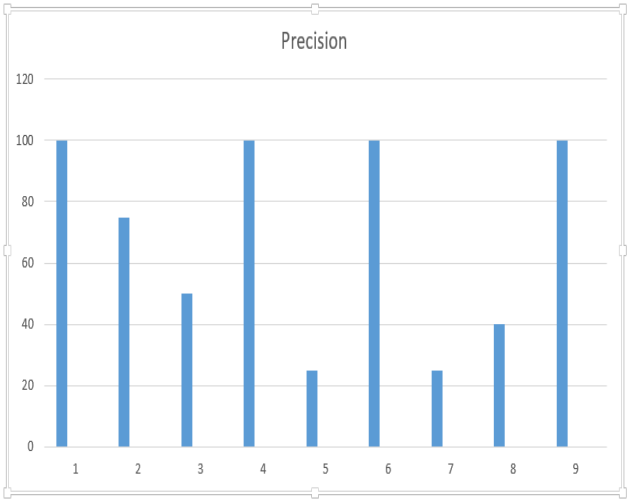

Figure 5: Average precision of the similarity search method

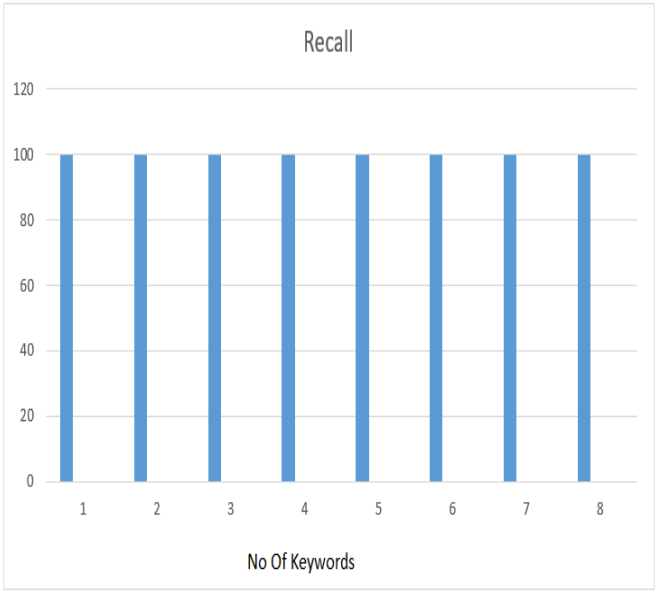

Figure 6: Average Recall of the similarity search method

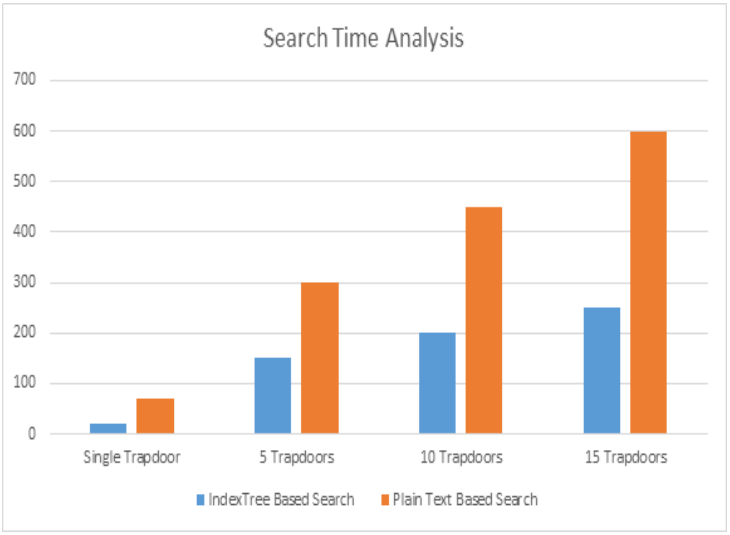

Figure 7: Search Time Analysis

VII CONCLUSION

In this System, a safe, effective and dynamic pursuit conspire is proposed, which underpins not just the precise multi-keyword positioned look yet in addition the dynamic cancellation and inclusion of archives. We develop a unique watchword adjusted paired tree as the record, and propose a "Ravenous Profundity initially Search" calculation to acquire better proficiency than direct hunt. Also, the parallel pursuit process can be done to additionally lessen the time cost. The security of the plan is ensured against two risk models by utilizing the protected KNN calculation. Trial comes about illustrate the effectiveness of our proposed conspire. There are as yet many test issues in symmetric SE plans. In the proposed plot, the information proprietor is dependable for producing refreshing data and sending them to the cloud server. Along these lines, the information proprietor needs to store the decoded list tree and the data that are important to recalculate the IDF esteems. Such a dynamic information proprietor may not be extremely reasonable for the distributed computing model. It could be an important however troublesome future work to outline a dynamic accessible encryption plot whose refreshing operation can be finished by cloud server just, in the interim saving the capacity to help multi-watchword positioned look. Furthermore, as the a large portion of works about accessible encryption, our plan for the most part considers the test from the cloud server. In reality, there are many secure difficulties in a multi-client conspire. To start with, every one of the clients for the most part keep the same secure key for trapdoor age in a symmetric SE plot. For this situation, the repudiation of the client is huge test. On the off chance that it is expected to repudiate a client in this plan, we require to remake the record

and disseminate the new secure keys to all the approved clients. Second, symmetric SE plots as a rule accept that every one of the information clients are dependable. It isn't reasonable what's more, an untrustworthy information client will prompt many secure issues. For instance, an untrustworthy information client may look through the records and convey the unscrambled archives to the unapproved ones. Considerably more, an unscrupulous information client may circulate his/her safe keys to the unapproved ones. In the future works, we will endeavor to enhance the SE plan to handle these test issues.

REFERENCES

1. B. Wang, S. Yu, W. Lou, and Y. T. Hou, " Privacy-preserving multi-keyword fuzzy search over encrypted data in the cloud ", in Proc.IEEE INFOCOM, 2014.
2. W. Zhang, S. Xiao, Y. Lin, T. Zhou, and S. Zhou, " Secure ranked multi-keyword search for multiple data owners in cloud computing ", in Dependable Syst. Net-works (DSN), IEEE 44th Annu. IEEE/IFIP Int. Conf., 2014.
3. 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. Inf., Comput. Commun. secur., 2013.
4. C. Orencik, M. Kantarcioglu, and E. Savas, " A practical and secure multi-keyword search method over encrypted cloud data ", in Proc. IEEE 6th Int. Conf. Cloud Comput., 2013.
5. S. Kamara and C. Papamanthou, " Parallel and dynamic searchable symmetric encryption ", in Proc. Financ. Cryptography Data Secur., 2013.
6. D. Cash, S. Jarecki, C. Jutla, H. Krawczyk, M.-C. Rosu, and M. Steiner, " Highly-scalable searchable symmetric encryption with support for boolean queries ", in Proc. Adv. Cryptol, 2013.
7. C. Wang, N. Cao, K. Ren, and W. Lou, " Enabling secure and efficient ranked keyword search over outsourced cloud data ", IEEE Trans. Parallel Distrib. Syst., vol. 23, no. 8, pp. 14671479, Aug. 2012.
8. C. Wang, K. Ren, S. Yu, and K. M. R. Urs, " Achieving usable and privacy-assured similarity search over outsourced cloud data ", in Proc. IEEE INFO-COM, 2012.
9. M. Kuzu, M. S. Islam, and M. Kantarcioglu, " Efficient similarity search over encrypted data ", in Proc. IEEE 28th Int. Conf. Data Eng., 2012.
10. S. Kamara, C. Papamanthou, and T. Roeder, " Dynamic searchable symmetric encryption ", in Proc. ACM Conf. Comput. Commun. Secur., 2012.
11. B. Zhang and F. Zhang, " An efficient public key encryption with conjunctive-subset keywords search ", J. Netw. Comput. Appl., vol. 34, no. 1, pp. 262267, 2011.
12. N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, " Privacy-preserving multi-keyword ranked search over encrypted cloud data ", in Proc. IEEE INFOCOM, Apr. 2011.