

A Secure and Dynamic Multi-Keyword Ranked Search Scheme over Encrypted Cloud Data

Vishesh J Sangshetty¹ Dr. H D Phaneendra²

¹PG Scholar ²Professor and HOD

^{1,2}Department of Computer Science and Engineering

^{1,2}National Institute of Engineering (NIE), Mysuru, Karnataka, India

Abstract— Due to the growing recognition of cloud computing, increasingly facts proprietors are encouraged to outsource their statistics to cloud servers for notable convenience and decreased price in facts management. However, touchy facts must be encrypted before outsourcing for privacy requirements, which obsoletes facts utilization like keyword-based file retrieval. In this paper, we present a relaxed multi-keyword ranked search scheme over encrypted cloud records, which simultaneously support dynamic replace operations like deletion and insertion of documents. Specifically, the vector space model and the broadly-used TF IDF version are blended within the index creation and question era. We construct a unique tree-based index structure and suggest a “Greedy Depth-first Search” set of rules to offer efficient multi-key-word ranked seek. The comfy kNN algorithm is applied to encrypt the index and question vectors, and meanwhile ensure accurate relevance score calculation between encrypted index and question vectors. In order to face up to statistical attacks, phantom phrases are delivered to the index vector for blinding search outcomes. Due to using our unique tree-primarily based index Structure, the proposed scheme can acquire sub-linear seek time and deal with the deletion and insertion of files flexibly. Extensive experiments are carried out to demonstrate the efficiency of the proposed scheme.

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

I. INTRODUCTION

Cloud computing has been considered as a brand new model of enterprise IT infrastructure, that could organize massive aid of computing, storage and programs, and enable customers to experience ubiquitous, handy and on-demand community get right of entry to to a shared pool of configurable computing assets with notable performance and minimum economic overhead. Attracted by means of these appealing functions, both people and businesses are inspired to outsource their statistics to the cloud, instead of buying software and hardware to manage the facts themselves.

Despite of the diverse benefits of cloud services, outsourcing sensitive facts (consisting of e-mails, private health information, organization finance data, authority’s files, and many others.) to far fling servers brings privations issues. The cloud carrier carriers (CSPs) that preserve the records for customers may get right of entry to customers’ touchy facts without authorization. A standard technique to shield the records confidentiality is to encrypt the records earlier than outsourcing. However, this could reason a massive value in phrases of information usability.

For example, the present strategies on keyword-based records retrieval, which might be broadly used at the plaintext records, can’t be without delay implemented on the

encrypted records. Downloading all of the records from the cloud and decrypt domestically is obviously impractical. In order to deal with the above trouble, researchers have designed a few preferred-reason answers with fully-homomorphic encryption or oblivious RAMs.

However, these methods are not realistic because of their excessive computational overhead for both the cloud sever and consumer. On the contrary, extra realistic special reason solutions, along with searchable encryption (SE) schemes have made unique contributions in terms of performance, capability and protection. Searchable encryption schemes enable the customer to keep the encrypted statistics to the cloud and execute keyword seek over cipher text domain. So some distance, ample works had been proposed beneath unique danger models to achieve numerous seek capability, which includes single key-word seek, similarity seek, multi-key-word Boolean seek, ranked seek, multi-key-word ranked search, and so on.

Among them, multi keyword ranked seek achieves increasingly more interest for its sensible applicability. Recently, a few dynamic schemes had been proposed to help putting and deleting operations on file collection. These are extensive works as it’s far exceedingly feasible that the statistics proprietors need to update their facts at the cloud server. But few of the dynamic schemes guide green multi key-word ranked seek.

II. RELATED WORK

The encrypted data to the cloud and execute keyword search over cipher text domain. Due to different cryptography Primitives, searchable encryption schemes can be constructed using public key based cryptography. or symmetric key based cryptography. Song *et al.* proposed the first symmetric searchable encryption (SSE) scheme, and the search time of their scheme is linear to the size of the data collection. Goh [8] proposed formal security definitions for SSE and designed a scheme based on Bloom filter. The search time of Goh’s scheme is $O(n)$, where n is the cardinality of the document collection. Curtmola *et al.* [10] proposed two schemes (SSE-1 and SSE-2) which achieve the optimal search time. Their SSE-1 scheme is secure against chosen-keyword attacks (CKA1) and SSE-2 is secure against adaptive chosen-keyword attacks (CKA2). These early works are single keyword Boolean search schemes, which are very simple in terms of functionality. Afterward, abundant works have been proposed under different threat models to achieve various search functionality, such as single keyword search, similarity, multi-keyword Boolean search, ranked search, and multi-keyword ranked search etc.

Multi-keyword Boolean search allows the users to input multiple query keywords to request suitable documents. Among these works, conjunctive keyword search schemes

only return the documents that contain all of the query keywords. Disjunctive keyword search schemes return all of the documents that contain a subset of the query keywords. Predicate search schemes are proposed to support both conjunctive and disjunctive search. All these multikeyword search schemes retrieve search results based on the existence of keywords, which cannot provide acceptable result ranking functionality. 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. Cao *et al.* realized the first privacy-preserving multi-keyword ranked search scheme, in which documents and queries are represented as vectors of dictionary size. With the “coordinate matching”, the documents are ranked according to the number of matched query keywords. However, Cao *et al.*'s scheme does not consider the importance of the different keywords, and thus is not accurate enough. In addition, the search efficiency of the scheme is linear with the cardinality of document collection.

Sun *et al.* presented a secure multi-keyword search scheme that supports similarity-based ranking. The authors constructed a searchable index tree based on vector space model and adopted cosine measure together with $TF \times IDF$ to provide ranking results. Sun *et al.*'s search algorithm achieves better-than-linear search efficiency but results in precision loss. O' rencik *et al.* proposed a secure multi-keyword search method which utilized local sensitive hash (LSH) functions to cluster the similar documents. The LSH algorithm is suitable for similar search but cannot provide exact ranking. In , Zhang *et al.* proposed a scheme to deal with secure multi-keyword ranked search in a multi-owner model. In this scheme, different data owners use different secret keys to encrypt their documents and keywords while authorized data users can query without knowing keys of these different data owners. The authors proposed an “Additive Order Preserving Function” to retrieve the most relevant search results. However, these works don't support dynamic operations.

III. PROBLEM STATEMENT

A. Existing Model:

A trendy method to protect the facts confidentiality is to encrypt the information earlier than outsourcing. Searchable encryption schemes allow the client to shop the encrypted data to the cloud and execute keyword search over cipher text area. So far, plentiful works were proposed under unique hazard models to acquire numerous seek capability, together with unmarried key-word search, similarity search, multi-keyword Boolean seek, ranked seek, multi-keyword ranked seek, and many others. Among them, multi-key-word ranked search achieves more and more interest for its practical applicability. Recently, a few dynamic schemes were proposed to aid inserting and deleting operations on report series. These are large works as it's miles distinctly feasible that the information owners want to update their information on the cloud server.

Disadvantages:

- Huge cost in terms of records usability. For instance, the prevailing strategies on key-word-based information retrieval, which can be widely used at the plain text records, can't be directly carried out at the encrypted statistics. Downloading all of the records from the cloud and decrypt domestically is obviously impractical.
- Existing System techniques not practical due to their high computational overhead for each the cloud sever and person.

B. Proposed Model:

- This paper proposes a relaxed tree-primarily based seek scheme over the encrypted cloud facts, which helps multi-keyword ranked seek and dynamic operation at the document series. Specifically, the vector area version and the widely-used “term frequency (TF) \times inverse file frequency (IDF)” model are combined within the index construction and question generation to provide multi-keyword ranked search. In order to attain high seek performance, we construct a tree-based totally index structure and endorse a “Greedy Depth-first Search” set of rules based on this index tree.
- The comfy kNN algorithm is utilized to encrypt the index and query vectors, and in the meantime ensure accurate relevance score calculation between encrypted index and question vectors.
- To face up to exceptional attacks in specific chance fashions, we assemble secure seek schemes: the basic dynamic multi-keyword ranked search (BDMRS) scheme in the recognized cipher textual content version, and the improved dynamic multi-key-word ranked seek (EDMRS) scheme in the recognized heritage model.

Advantages:

Due to the special shape of our tree-based totally index, the proposed seek scheme can flexibly achieve sub-linear seek time and cope with the deletion and insertion of documents.

We design a searchable encryption scheme that helps both the accurate multi-key-word ranked seek and flexible dynamic operation on document series.

Due to the special structure of our tree-based index, the hunt complexity of the proposed scheme is essentially stored to logarithmic. And in exercise, the proposed scheme can attain better search efficiency with the aid of executing our “Greedy Depth-first Search” algorithm. Moreover, parallel search may be flexibly executed to similarly reduce the time value of seek process.

IV. PROBLEM FORMULATION

A. Notations & Preliminaries:

- \mathcal{W} —The dictionary, namely, the set of keywords, denoted as $\mathcal{W} = \{w_1, w_2, \dots, w_m\}$.
- m —The total number of keywords in \mathcal{W} .
- \mathcal{W}_q —The subset of \mathcal{W} , representing the keywords in the query.
- \mathcal{F} —The plaintext document collection, denoted as a collection of n documents $\mathcal{F} = \{f_1, f_2, \dots, f_n\}$. Each document f in the collection can be considered as a sequence of keywords.
- n —The total number of documents in \mathcal{F} .
- \mathcal{C} —The encrypted document collection stored in the cloud server, denoted as $\mathcal{C} = \{c_1, c_2, \dots, c_n\}$.
- \mathcal{T} —The unencrypted form of index tree for the whole document collection \mathcal{F} .
- \mathcal{I} —The searchable encrypted tree index generated from \mathcal{T} .
- Q —The query vector for keyword set \mathcal{W}_q .
- TD —The encrypted form of Q , which is named as trapdoor for the search request.
- D_u —The index vector stored in tree node u whose dimension equals to the cardinality of the dictionary \mathcal{W} . Note that the node u can be either a leaf node or an internal node of the tree.
- I_u —The encrypted form of D_u .

The system model in this paper entails 3 exclusive Entities: data owner, data user and cloud server, As illustrated in Fig. 1.

Data owner has a group of documents $F = \{f_1; f_2; \dots; f_n\}$ that he wants to outsource to the cloud server in encrypted shape at the same time as still keeping the capability to go looking on them for powerful usage. In our scheme, the records owner firstly builds a comfy searchable tree index I from document series F , and then generates an encrypted do series C for F . Afterwards, the records proprietor outsources the encrypted collection C and the relaxed index I to the cloud server, and securely distributes the key information of trapdoor era (inclusive of key-word IDF values) and file decryption to the authorized data customers. Besides, the data Proprietor is accountable for the update operation of his files stored inside the cloud server. While updating, the facts owner generates the replace information regionally and sends it to the server.

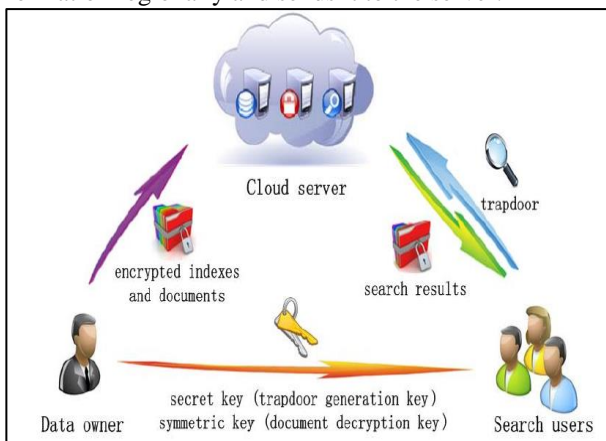


Fig. 1: System Architecture

Data customers are authorized ones to get entry to the documents of data proprietor. With t question keywords, the legal user can generate a trapdoor TD in step with seek manipulate mechanisms to fetch k encrypted files from cloud server. Then, the records user can decrypt the files with the shared mystery key.

Cloud server shops the encrypted record collection C and the encrypted searchable tree index I for statistics proprietor. Upon receiving the trapdoor TD from the data user, the cloud server executes seek over the index tree I , and sooner or later returns the corresponding collection of pinnacle- okay ranked encrypted files. Besides, upon receiving the replace information from the records proprietor, the server desires to replace the index I and report collection C according to the obtained data.

V. DESIGN GOALS

To enable secure, efficient, accurate and dynamic multi data under the above models, our system has the following

Dynamic: The proposed scheme is designed to provide not only multi-keyword query and accurate result rating, however also dynamic replace on record collections.

Search Efficiency: The scheme aims to obtain sub linear seek performance by way of exploring a special tree-based totally index and an green search set of rules.

A. Privacy-maintaining:

The scheme is designed to prevent the cloud server from learning additional facts approximately the document series, the index tree, and the question. The unique privacy requirements are summarized as follows,

B. Index Confidentiality and Query Confidentiality:

The underlying undeniable textual content data, such as keywords within the index and question, TF values of keywords stored in the index, and IDF values of query keywords, must be covered from cloud server;

C. Trapdoor Unlink ability:

The cloud server need to not be able to determine whether encrypted queries (trapdoors) are generated from the equal seek request;

D. Keyword Privacy:

The cloud server could not become aware of the unique key-word in question, index or report series by using analyzing the statistical statistics like time period frequency. Note that our proposed scheme isn't designed to guard get entry to sample, i.e., the sequence of lower back documents.

VI. CONCLUSION

In this paper, a secure, efficient and dynamic search scheme is proposed, which helps no longer most effective the correct multi key-word ranked seek but also the dynamic deletion and insertion of documents. We construct a unique key-word balanced binary tree because the index, and advice a "Greedy Depth-first Search" set of rules to obtain higher efficiency than linear search. In addition, the parallel seek system may be carried out to similarly lessen the time value. The security of the scheme is blanketed in opposition to threat fashions through using the at ease kNN algorithm. Experimental consequences exhibit the performance of our proposed

scheme. There are nevertheless many mission troubles in symmetric SE schemes. In the proposed scheme, the records proprietor is responsible for generating updating information and sending them to the cloud server. Thus, the data owner wishes to store the unencrypted index tree and the facts which are essential to recalculate the IDF values. Such an energetic records proprietor might not be very suitable for the cloud computing model. It may be a significant but difficult destiny work to design a dynamic searchable encryption scheme whose updating operation may be completed by using cloud server only, meanwhile reserving the capability to aid multi-key-word ranked search. In addition, because the most of works approximately searchable encryption, our scheme specially considers the project from the cloud server. Actually, there are many comfortable demanding situations in a multi-consumer scheme. First, all of the users normally preserve the equal secure key for trapdoor era in a symmetric SE scheme. In this situation, the revocation of the person is big project. If it is needed to revoke a person in this scheme, we need to rebuild the index and distribute the brand new comfortable keys to all the legal users. Second, symmetric SE schemes commonly expect that all the records users are truthful. It isn't always sensible and a unethical statistics consumer will result in many at ease issues. For instance, a unethical statistics user may seek the documents and distribute the decrypted files to the unauthorized ones. Even greater, a dishonest records user can also distribute his/her comfortable keys to the unauthorized ones. In the destiny works, we will attempt to improve the SE scheme to deal with these project problems.

REFERENCES

- [1] K. Ren, C. Wang, Q. Wang et al., "Security challenges for the public cloud," *IEEE Internet Computing*, vol. 16, no. 1, pp. 69–73, 2012.
- [2] S. Kamara and K. Lauter, "Cryptographic cloud storage," in *Financial Cryptography and Data Security*. Springer, 2010, pp. 136–149.
- [3] C. Gentry, "A fully homomorphic encryption scheme," Ph.D. dissertation, Stanford University, 2009.
- [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. Di Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in *Advances in Cryptology- Eurocrypt 2004*. Springer, 2004, pp. 506–522.
- [6] D. Boneh, E. Kushilevitz, R. Ostrovsky, and W. E. Skeith III, "Public key encryption that allows pir queries," in *Advances in Cryptology-CRYPTO 2007*. Springer, 2007, pp. 50–67.
- [7] D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in *Security and Privacy, 2000. S&P 2000. Proceedings. 2000 IEEE Symposium on*. IEEE, 2000, pp. 44–55.
- [8] E.-J. Goh et al., "Secure indexes." *IACR Cryptology ePrint Archive*, vol. 2003, p. 216, 2003.
- [9] Y.-C. Chang and M. Mitzenmacher, "Privacy preserving keyword searches on remote encrypted data," in *Proceedings of the Third international conference on Applied Cryptography and Network Security*. Springer-Verlag, 2005, pp. 442–455.
- [10] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: improved definitions and efficient constructions," in *Proceedings of the 13th ACM conference on Computer and communications security*. ACM, 2006, pp. 79–88.
- [11] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, "Fuzzy keyword search over encrypted data in cloud computing," in *INFOCOM, 2010 Proceedings IEEE*. IEEE, 2010, pp. 1–5.
- [12] M. Kuzu, M. S. Islam, and M. Kantarcioglu, "Efficient similarity search over encrypted data," in *Data Engineering (ICDE), 2012 IEEE 28th International Conference on*. IEEE, 2012, pp. 1156–1167.