International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

OTP Product resemblance Search Scheme for Encrypted Cloud Data

A. Ramaswami Reddy

Assistant Professor, Computer Science Engineering, Vignan's Foundation for Science, Technology & Research (Deemed to be University) Deemed university in Guntur, Andhra Pradesh

Abstract:

A Secure and Dynamic Multi-keyword Ranked Search Scheme over Encrypted Cloud Data Due to the increasing popularity of cloud computing, more and more data owners are motivated to outsource their data to cloud servers for great convenience and reduced cost in data management. However, sensitive data should be encrypted before outsourcing for privacy requirements, which obsoletes data utilization like keyword-based document retrieval. In this application, we present a secure multi-keyword ranked search scheme over encrypted cloud data, which simultaneously supports dynamic update operations like deletion and insertion of documents. Specifically, the vector space model and the widely-used TFIDF model are combined in the index construction and query generation. 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 vectors. In order to resist statistical attacks, phantom terms are added to the index vector for blinding search results. Due to the use of our special tree-based index structure, the proposed scheme can achieve sub-linear search time and deal with the deletion and insertion of documents flexibly. Extensive experiments are conducted to demonstrate the efficiency of the proposed scheme.

Keywords— Multi-keyword ranked search over encrypted cloud data, OTP, Product resemblance, Cloud, Data owners

1. Introduction

Now a day's cloud computing has become essential for many utilities, where cloud customers can slightly store their data into the cloud so as to benefit from on-demand high-quality request and services from a shared pool of configurable computing resources. Its huge suppleness and financial savings are attracting both persons and enterprise to outsource their local complex data management system into the cloud. To safe guard data privacy and struggleunwanted accesses in the cloud and away from, sensitive data, for example, emails, personal health records, photo albums, videos, land documents, financial transactions, and so on, may have to be encrypted by data holder before outsourcing to the business public cloud; on the other hand, obsoletes the traditional data use service based on plaintext keyword search. The insignificant solution of downloading all the information and decrypting nearby is clearly impossible, due to the enormous amount of bandwidth cost in cloud scale systems. Furthermore, apart from eradicating the local storage management, storing data into the cloud supplies no purpose except they can be simply searched and operated. Thus, discovering privacy preserving and effective search service over encrypted cloud data is one of the supreme importance. In view of the potentially large number of on-demand data users and vast amount of outsourced data documents in the cloud, this difficulty is mostly demanding as it is really difficult to gather the

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

requirements of performance, system usability, and scalability.

On the one hand, to congregate the efficient data retrieval requirement, the huge amount of documents orders the cloud server to achieve result relevance ranking, as an alternative of returning undifferentiated results. Such ranked search system allows data users to discover the most appropriate information quickly, rather than burdensomely sorting during every match in the content group. Ranked search can also gracefully remove redundant network traffic by transferring the most relevant data, which is highly attractive in the "pay-as-you-use" cloud concept. For privacy protection, such ranking operation on the other hand, should not reveal any keyword to related information. To get better the search result exactness as well as to improve the user searching experience, it is also essential for such ranking system to support multiple keywords search, as single keyword search often give up far too common results. As a regular practice specifies by today's web search engines i,e Google search, data users may lean to offer a set of keywords as an alternative of only one as the indicator of their search interest to retrieve the most relevant data. And each keyword in the search demand is able to help narrow down the search result further. "Coordinate matching", as many matches as possible, is an efficient resemblance measure among such multi-information retrieval (IR) community. Though, the nature of applying encrypted cloud data search system remains a very demanding task in providing security and maintaining privacy, like the data privacy, the index privacy, the keyword privacy, and many others. Encryption is a helpful method that treats encrypted data as documents and allows a user to securely search through a single keyword and get back documents of interest. On the other hand, direct application of these approaches to the secure large scale cloud data utilization system would not be necessarily suitable, as they are developed as crypto primitives and cannot put up such high service-level needs like system usability, user searching experience, and easy information discovery. Even though some modern plans have been proposed to carry Boolean keyword search as an effort to improve the search flexibility, they are still not sufficient to provide users with satisfactory result ranking functionality. The solution for this problem is to secure ranked search over encrypted data but only for queries consisting of a single keyword. The challenging issue here is how to propose an efficient encrypted data search method that supports multi-keyword semantics without privacy violation.

2. RELATED WORK

Traditional searchable encryption has been widely studied as a cryptographic primitive, with a focus on security definition formalizations and efficiency improvements. Song et al. first introduced the notion of searchable encryption. They proposed a scheme in the symmetric key setting, where each word in the file is encrypted independently under a special two-layered encryption construction. Thus, a searching overhead is linear to the whole file collection length. Goh developed a Bloom filter-based per-file index, reducing the workload for each search request proportional to the number of files in the collection. Chang and Mitzenmacher also developed a similar per-file index scheme. To further enhance search efficiency, Curtmola et al. proposed a per-keyword-based approach, where a single encrypted hash table index is built for the entire file collection, with each entry consisting of the trapdoor of a keyword and an encrypted set of related file identifiers. Searchable encryption has

also been considered in the public-key setting. Aiming at tolerance of both minor typos and format inconsistencies in the user search input, fuzzy keyword search over encrypted cloud data has been proposed by Li et al. in [9]. Very recently, a privacy-assured similarity search mechanism over outsourced cloud data has been explored by Wang et al. in [2]. Note that all these schemes support only Boolean keyword search and none of them support the ranked search problem which we are focusing on in this paper. Following our research on secure ranked search over encrypted data, very recently, Cao et al. [1] propose a privacy-preserving multi keyword ranked search scheme, which extends our previous work in [1] with support of multi keyword query. They choose the principle of "coordinate matching," i.e., as many matches as possible, to capture the similarity between a multi keyword search query and data documents and later quantitatively formalize the principle by a secure inner product computation mechanism. One disadvantage of the scheme is that cloud server has to linearly traverse the whole index of all the documents for each search request, while ours is as efficient as existing SSE schemes with only constant search cost on cloud server.

Secure top-k retrieval from Database Community from database community are the most related work to our proposed RSSE. The idea of uniformly distributing posting elements using an order-preserving cryptographic function. The order preserving mapping function proposed does not support score dynamics, i.e., any insertion and updates of the scores in the index will result in the posting list completely rebuilt. Zerr et al. use a different order-preserving mapping based on presampling and training of the relevance scores to be outsourced, which is not as efficient as our proposed schemes. Besides, when scores following different distributions need to be inserted, their score transformation function still needs to be rebuilt. On the contrary, in our scheme the score dynamics can be gracefully handled, which is an important benefit inherited from the original OPSE. This can be observed from the Binary Search (.). In other words, the newly changed scores will not affect previous mapped values. We note that supporting score dynamics, which can save quite a lot of computation overhead when file collection changes, is a significant advantage in our scheme. Moreover, both works above do not exhibit thorough security analysis which we do in the paper.

3. Proposed Multi-Keyword Ranked Search Over Encrypted (PMRSE)

In this paper, we describe and solve the problem of multi-keyword ranked search over encrypted cloud data (PMRSE) while preserving exact system wise privacy in the cloud computing concept. Along with various multi-keyword semantics, select the efficient resemblance measure of "coordinate matching," it means that as various matches as possible, to confine the significance of data documents to the search query. Particularly, inner product similarity the numbers of query keywords show in a document, to quantitatively calculate such similarity assess of that document to the search query. For the period of the index construction, each document is associated with a binary vector as a sub-index where each bit signifies whether matching keyword is contained in the document. The search query is also illustrates as a binary vector where each bit means whether corresponding keyword appears in this search request, so the resemblance could be exactly calculated by the inner product of the query vector with the data vector. On the other hand, directly outsourcing the data vector or the

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

queryvector will go against the index privacy or the search privacy. To face the challenge of cooperating such multi keyword semantic without privacy breaches, we propose a basic idea for the MRSE using secure inner product computation, which is modified from a secure k-nearest neighbour (kNN) method, and then give two considerably improved MRSE method in a step-by-step way to accomplish different severe privacy needs in two risk models with enlarged attack competence.

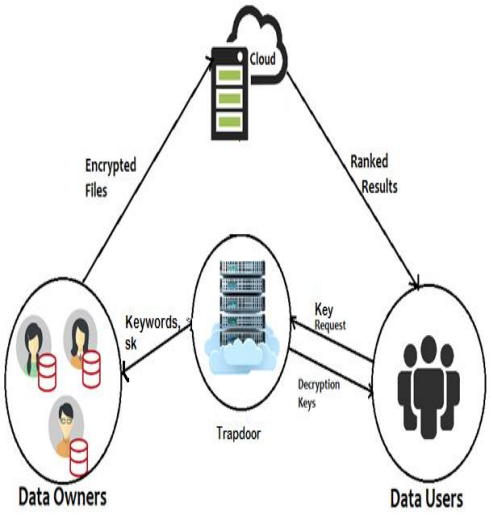

Fig 1.Architecture

Solution Architecture

Data sources

Data sources for this tool comprises of various data warehouse tables like transactional data, listing data, Behavioral data, user data and owners to user linking data. Owner upload data, can be stored in database, and user can search data from database.

Segmentation platform

Segmentation platform is the frontend application that will be used to define the segmentation models by the analysts using the even driven approach of guiding the analyst with different options for the segmentation. Segmentation metadata will store the segment information, segmented member information and segmentation thresholds entered by the analyst. SQLs for the segmentation models could be generated and exported using this application.

4. PROPOSED METHODOLOGY

4.1 Framework

The framework will be present a searchable encryption scheme that supports both the accurate multi-keyword ranked search and flexible dynamic operation on document collection. This framwork proposes a secure tree-based search scheme over the encrypted cloud data, which supports multikeyword 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 multikeyword ranked search.

1. Abundant works have been proposed under different threat models to achieve various search functionality,
2. Recently, some dynamic schemes have been proposed to support inserting and deleting operations on document collection.

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.

Algorithm: Term Frequency-Inverse Document Frequency

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

Input: *Data d.*

Output: *result r.*

Let data d,

Collection c;

c=getWords(d); //Using Split("\\s+")

Term Frequency tf;

α= Number of times term t appears in a document;

tf=(α);.

Inverse Document Frequency idf;

α= Number of times term t appears in a document;

β= Total number of terms in the document;

IDF(t) = (α)/(β);.

End;

4.2 Hidden Keyword Tree Structure

Here we propose hidden keyword tree structure formed by keyword searchable cipher texts. The user can delegate the keyword search to the server. More specifically, each owner separately encrypts a file and its extracted keywords and sends the resulting cipher texts to a server; and those form like tree structure data like fig. Every file id will connect to tree leaf when same match is found otherwise create another leaf. When the user wants to retrieve the files containing a specific keyword, he/she delegates a keyword search trapdoor to the server; the server finds the encrypted files containing the queried keyword without knowing the original files or the keyword itself, and returns the corresponding encrypted files to the user; finally, the receiver decrypts these encrypted files.

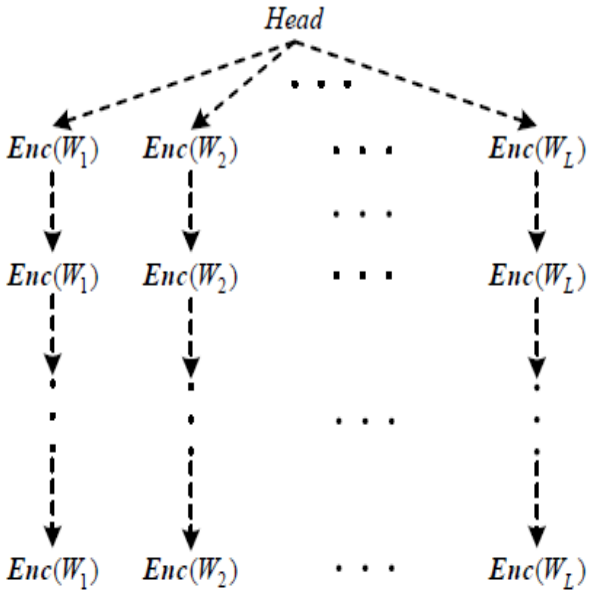

Fig 2. Hidden Tree Structure

5. RESULTS AND DISCUSSION

The proposed scheme, data users can achieve different requirements on search precision of privacy by the standard deviation of adjustment that can be treated as a compensation parameter. The comparison of systems with a recent work that achieves high search efficiency. PMRSEscheme calls the search results by exact calculation of document vector and query vector. Thus, top- k search accuracy of PMRSEscheme is 100 %. But based and similarity Multi- keyword square search pattern, the basic scheme in suffering from loss of precision due to the accumulation of sub-vectors with the index construction . The test is repeated 16 times, and the average accuracy of 91 %. During the search, when the relevance of the node is greater than the minimum relevance in results Rlist, examines the cloud server. Otherwise it returns. So many nodes not accessed during a real search. We denote the number of leaf nodes that contain one or more keywords in the query. It is generally greater than the number of documents required k, but far

less than the cardinality of the document collection n. As a balanced binary tree, the height of the index n is log will be maintained, and the complexity of the calculation is ranked relevance O (m).

6. CONCLUSION AND FUTURE WORK

In this paper we describe and solve the problem of multikey word ranked search over encrypted cloud data, and set up a range of privacy requirements. Among various multi-keyword semantics, we select the efficient similarity measure of "coordinate matching," i.e., as many equivalent as possible, to effectively capture the relevance of outsourced documents to the query Keywords, and utilize "inner product similarity" to quantitatively calculate such comparison measure. In order to acquire the test of supporting multi-keyword semantic without privacy violation, we offer a basic idea of MRSE using secure inner product calculation. Then, we give two improved MRSE schemes to attain various severe privacy needs in two different threat models. The further enhancements of our ranked search method, including supporting more search semantics, i.e., TF _ IDF, and dynamic data process. Detailed analyses in investigating privacy and efficiency assurance of proposed schemes are mentioned, and testing on the real-world data set demonstrate our proposed schemes which introduces low transparency on both calculation and communication.

Scope for Future Extension

With the advent of cloud computing, it has become increasingly popular for data owners to outsource their data to public cloud servers while allowing data users to retrieve this data. For privacy concerns, secure searches over encrypted cloud data have motivated several research works under the single owner model. However, most cloud servers in practice do not just serve one owner; instead, they support multiple owners to share the benefits brought by cloud computing. We enhancement schemes to deal with Privacy preserving Ranked Multi-keyword Search in a Multi-owner model. To enable cloud servers to perform secure search without knowing the actual data of both keywords and trapdoors, we systematically construct a novel secure search protocol. To reduce the computation cost trapdoors are not invalid in owns side to generate keywords.

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