A Survey On Ranked Multikeyword Search Over Outsourced Cloud Data

Lekshmi Balakrishnan, Soja Salim

M.Tech Student, Dept. of CSE, SCTCE, Pappanamcode, Trivandrum, India;
Assistant Professor, Dept. of CSE, SCTCE, Pappanamcode, Trivandrum, India.
Email: leksh123bala@gmail.com

ABSTRACT: With the arrival of cloud computing, data owners can outsource their complex data from local systems to the commercial public cloud .Storing data on cloud enables them to access their data from anywhere by using any system. The main advantage of cloud computing is data service outsourcing.It allows data owners to store their documents in the public data centers by economically saving their capital investment towards data management. But for protecting data privacy, sensitive information must be encrypted before storing them to cloud.Encrypting the documents before storing them to cloud outmoded traditional data utilization which focuses on plaintext keyword search. Today's main challenge is to enable searching over encrypted cloud data. Since cloud contains large number of data users and documents, it is necessary to allow multi-keywords in the search request and return documents in the order of their relevance to these keywords. The main aim is to make effective data access in the public cloud storage by improving various searching techniques, thereby, increasing the data utilization. In this survey, an attempt is made to study various searching techniques over encrypted cloud data.

Keywords: Cloud computing; searchable encryption;privacy-preserving;ranked search

1 Introduction

IN cloud computing, applications and files are stored on the "cloud" which consists of many computers and servers. All computers and servers are linked together and can be accessed through the Internet. Flexibility plays an important role in cloud computing which means the ability to meet the demand of users. Because of its great flexibility and economic savings both individuals and enterprises are allowed to outsource their data into the cloud. But, in order to protect the sensitive information, data owners must encrypt their data before outsourcing it to public cloud. Searching over the encrypted cloud data is a challenging task. Even though there are many traditional searchable encryption schemes which allows user to search over the encrypted data, most of them supports only exact keyword searching. A series of searchable symmetric encryption (SSE) schemes have been proposed which allows users to search on cipher text. Traditional schemes only allows users to securely retrieve the cipher text, but they support only Boolean keyword search, i.e., whether a keyword exists in a document or not. In order to improve security without losing data confidentiality, SSE schemes support top-k single keyword retrieval under various scenarios. For effective data retrieval, the cloud server performs ranked search over the encrypted data so that it returns the most relevant documents. Such ranked search system enables data users to find the most relevant information quickly, instead of burdensomely sorting through every match in the content collection .Ranked search also eliminate unnecessary network traffic because cloud server send back only the most relevant data, which is highly desirable in the "pay-as-you-use" cloud paradigm. These ranking operations should not leak any keyword related information for privacy protection. Also it is necessary for such ranking schemes to support multiple keywords in the search query to improve the search result accuracy. Coordinate matching is an efficient similarity measure among such multi-keyword semantics .It will return the result according to their relevance, However, how to apply it in the encrypted cloud data search system remains a very challenging task because of inherent security and privacy obstacles, including various strict requirements like the data privacy, the index privacy, the keyword privacy, etc.

2 Different Techniques To Search Over Encrypted Cloud Data

2.1 Similarity Search over Encrypted Data

Documents are encrypted before it is stored in the cloud so that only authorised users are allowed to access them. There are several techniques which allow cloud server to search over these encrypted data. But most of them handle only exact query matching. Efficient Similarity Search technique[1] not only handles exact query matching, but also matches query based on its similarity with documents. In this method, documents are retrieved if its similarity against a specified query word is greater than or equal to a predetermined threshold. To perform efficient search over the encrypted data, an index is constructed using Locality Sensitive Hashing (LSH). The locality sensitive hashing (LSH) uses hash functions to map similar objects into the same hash buckets with higher probability. Using the LSH index, it is easier to perform a similarity search. First using the LSH functions, matching documents are selected for a given query word. Then the selected documents are matched according to their similarity scores. . Those documents which have higher scores are returned to the user. In this method, confidentiality of the data is preserved. This method support efficient keyword search which is free from typographical errors. But this method has the disadvantage that it does not scale well when the number of data items is large. So this method does not work well for data stored in cloud.

2.2 Practical Techniques for Searches on Encrypted Data

"Practical techniques for searches on encrypted data" proposed by D. W. D. Song and A. Perrig[4], describes solution for the problem of searching on encrypted data. This paper introduces sequential scanning search technique .This technique searches over encrypted data in cloud without losing data confidentiality. The technique is provably secure in the sense that the cloud server doesn't know anything about the data stored in the cloud.It just stores all these data and isolates the query result so that the server doesn't know anything more than the search result. It also supports controlled searching by server ,i.e server is not allowed to search over the encrypted data without users request. These techniques perform probabilistic searching.It search for a word in the document collection and return back all the locations where the word appears.These techniques have a number of advantages such as 1) they are provably secure, 2) they support controlled and hidden search and query isolation, 3) they are simple and fast .Also these methods introduce no space and communication overhead. They are also flexible so that they can be extended to support search queries which are combined with Boolean operators, proximity queries, queries that contain regular expression, etc.But these techniques have disadvantage that , in case of large documents demanding huge volumes of storage, the technique has high time complexity.

2.3 Encrypted Phrase Searching in the Cloud

“Encrypted phrase searching in the cloud”proposed by S. Zittrower and Cliff C. Zou[5], provides a new method in the field of encrypted searching .It allows for both encrypted phrase searching and proximity ranked keyword searching over encrypted data stored in the cloud. There is no need to store documents or searchable index on local system. In this method, a trusted client-side server is used to encrypt and decrypt the files. The trusted client-side server generates the encrypted index from the document collection. After that both the encrypted documents and the searchable index is stored on the untrusted cloud. This method uses Proximity ranked searching which ranks documents by a function that is directly proportional to the distance between the keywords in the search phrase .The location information of keywords must be preserved in the encrypted index to perform proximity ranking. The user sends a query which contains single keyword or multikeyword to the trusted client-side server. The client-side server encrypts all keywords in the search query. It then truncates the encrypted keywords to a number of bits so as to improve security .Then the trusted client-side server sends the new query to the cloud server .The cloud server then searches the encrypted index for the keywords that matches the query. It returns the matching documents along with the encrypted index offset and keyword locations to the trusted client-side server. The trusted client-side server decrypts the data and then determines the actual documents and ranks them. Then it sends the ranked list to the user so that the user can retrieve the relevant document from the cloud server. The disadvantage of this method is that if the trusted server is compromised, the entire security is broken.

2.4 Toward Secure Multikeyword Top-k Retrieval over Encrypted Cloud Data

Jiadi[6] proposed a searching technique that uses Two- round searchable encryption (TRSE). In the first round, users submit an encrypted query to achieve privacy .The encrypted query may contain single or multiple keywords. A trapdoor is created for this query and this query is sent to the cloud server. The cloud server contains the encrypted documents and the index file. It computes the scores of documents from this index file and then returns the encrypted score result vector to user. In the second round, user decrypts the score with secret key and calculates the scores of files and then requests files with top-k scores. In this method the ranking is done on user side and computing score is done on server side. The TRSE utilizes vector space model and homomorphic encryption. The vector space model provides sufficient search accuracy, and the homomorphic encryption allows users to participate in the ranking. The majority of computing work is done on the server side .As a result, information leakage can be eliminated and data security is ensured. The user takes part in ranking, which guarantees top-k multikeyword retrieval over encrypted cloud data with high security and practical efficiency. This method finds a solution to the problem of secure multikeyword top-k retrieval over encrypted cloud data. But this method suffers from high communication overhead if the encrypted trapdoor size is too large.

2.5 Public-Key Encryption with Keyword Search (PEKS)

PEKS is said to be the first predicate encryption scheme. This was designed for the purpose of intelligent email routing. Soon this technique was enhanced to be integrated into the cloud environment. D. Boneh [7] proposed this search, in which cloud server contains encrypted files and keyword. User creates trapdoor using its private key to search a word. The cloud server checks the trapdoor with existing encrypted keyword and sends back the encrypted file that matches it. The owner stores the files in the cloud in encrypted format and server makes the user authentication, a secure channel exists between the owner, server and user. The four algorithms are used in this method. First, keyGen is used to generate public key and private key pair for both server and user. Second, PEKS algorithm produces searchable encryption. Third, Trapdoor algorithm is used to calculate trapdoor with private key and keyword. Fourth, Test is used to match the keyword and requested word. If it matches then the file is sent to the user. This scheme fails concerning access policy and dictionary attack. The major disadvantages are that trapdoor contains meaningful keywords and one-to-one mapping takes place between trapdoor and keyword. The disadvantage with this method is that, multiple keywords cannot be used for searching.

2.6 Multi-Keyword Ranked Search over Encrypted Cloud Data

In order to make data retrieval more effective, it is necessary to allow multiple keywords in the search request. Also to solve the challenging problem of privacy-preserving multi-keyword ranked search over encrypted data, this method (MRSE) [9] establishes a set of strict privacy requirements. Coordinate matching is used to capture the similarity between search query and documents. This method consists of four algorithms as follows:

1. Setup.
2. Build Index.
3. Trapdoor.
4. Query.

Setup phase uses KeyGen algorithm for generating public/private key. BuildIndex algorithm is used to generate index file which contains keywords from file. Both the encrypted file collection and index file is outsourced to cloud. When a user submits a query, based on some known parameters, a trapdoor is generated. After receiving the trapdoor, cloud server performs a search on the index file and then finally returns the ranked list, sorted by their relevance with the user’s keyword. This method chooses the efficient similarity measure of “coordinate matching” for capturing the relevance of outsourced documents with the query keywords and uses “inner product similarity” to quantitatively evaluate the similarity measure.This method proposes a basic idea of MRSE which uses secure inner product computation to support multikeyword searching over the encrypted data.

3 CONCLUSION

The survey on different techniques to search over the encrypted data solves the problem of ranked search over encrypted cloud data.All these methods allow users to perform keyword searching while improving the security of the user query.The cloud server performs searching over the encrypted data but server does not know the sensitive informations behind the data collection. Performing such kind of searching cause an increase in the computational cost and the cost associated with communication. Ranked search improves system performance by returning only the matching files in a ranked order. Various methods are used to protect the sensitive information from leaking. The main goal of all these methods is to prevent the cloud server from learning the sensitive information from the document set, the index file, and the user queries thus protecting user privacy. There still needs to find other methods so as to solve the multikeyword searching over the encrypted cloud data with least computation cost.

REFERENCES

[1] M. S. I. M. K. Mehmet Kuzu, “Efficient similarity search over encrypted data,” IEEE 28th International Conference on Data Engineering, 2012.

[2] P.Indyk and R.Motwani, “Approximate nearest neighbors: towards removing the curse of dimensionality,” in 30 th STOC, 1998.

[3] P. I. A. Gionis and R. Motwani., “A. gionis, p. indyk, and r. motwani.,” in Proc. of VLDB’99, 1999.

[4] D. W. D. Song and A. Perrig, “Practical techniques for searches on encrypted data,” in Proc. of S&P, 2000.

[5] S. Zittrower and C. C. Zou, “Encrypted phrase searching in the cloud,” in IEEE Symposium on Security and Privacy, 2012.

[6] Y. Z. G. X. J. Y. P. Lu and M. Li, “Toward secure multikeyword top-k retrieval over encrypted cloud ata,” IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, vol. 10, Aug 2013.

[7] D. Boneh and B. Waters, “Conjunctive, subset, and range queries on encrypted data,” Proc. Fourth Conf. Theory Cryptography (TCC),, pp. 535–554, 2007.

[8] E. S. E. Shen and B. Waters, “Predicate privacy in encryption systems,” Proc. Sixth Theory of Cryptography Conf. Theory of Cryptography (TCC), 2009.

[9] M. L. K. R. Ning Cao, Cong Wang and W. Lou, “Privacy-preserving multi-keyword ranked search over encrypted cloud data,” IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, 2014.