ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

A Survey on Secure and Aggressive Multi-Keyword Ranked Search Over Encrypted Cloud Data

Sucheta B. Patil [1], Mayank Bhatt [2]

M.Tech Scholar, Dept. of Computer Science & Engg, LNCTRIT, Indore, India [1]

Assistant Professor, Dept. of Computer Science & Engg, LNCTRIT, Indore, India [2]

ABSTRACT: As cloud computing is popular for motivated to outsource their data to cloud servers for great convenience and reduced cost in data management. This paper present a secure multi-keyword ranked search scheme over encrypted cloud data, which concurrently supports aggressive update operations like deletion and insertion of documents. In view of the huge number of data users and documents in the cloud, it is necessary to permit multiple keywords in the search request and return documents in the order of their relevance to these keywords. Due to the use of special tree-based Index structure, the proposed scheme can reach sub-linear search time and deal with the deletion and insertion of documents flexibly.

KEYWORDS: Multi-keyword ranked search, Aggressive update, Cloud computing

I. INTRODUCTION

Cloud computing has been considered as a new production of enterprise IT infrastructure, which can organize large resource of computing, storage and applications, and enable users to enjoy everywhere, convenient and on-demand network access of configurable computing resources with special efficiency and minimal economic overhead [1]. In spite of the various advantages of cloud services, outsourcing sentient information (such as e-mails, personal health records, company finance data, government documents, etc.) to remote servers shows privacy concerns. The cloud service providers (CSPs) that keep the data for users may access user's sentient information without permission. A general approach to preserve the data confidentiality is to encrypt the data before outsourcing [2]. However, this will cause a huge cost as to data usability. For example, the existing techniques on keyword-based information retrieval, which are mostly used on the plaintext data, that cannot be directly applied on the encrypted data. In cloud computing, scalable and elastic storage and computation resources are provisioned as measured services through the Internet. Outsourcing data services to the cloud allows organizations to be keep on not only monetary savings, but also simplified local IT management since cloud infrastructures are physically hosted and maintained by the cloud providers. To reduce the risk of data leakage to the cloud service providers, data owners opt to encrypt their sensitive data, e.g., health records, financial transactions, before outsourcing to the cloud, while retaining the decryption keys to themselves and other authorized users. This in turn renders data utilization a challenging problem. For example, in order to search some applicable documents amongst an encrypted data set stored in the cloud, one may have to download and decrypt the entire data set. This is evidently impractical when the data volume is large. Thus, mechanisms that allow users to search directly on the encrypted data are of more interest in the cloud computing era.[2]. In order to address the above problem, researchers have built some general-purpose solutions. Searchable encryption schemes specify the client to store the encrypted data to the cloud and execute keyword search over ciphertext domain. Data encryption makes powerful data utilization a very challenging taskgiven that there could be a large amount of outsourced data files. Except, in Cloud Computing, data owners may share their outsourced data with a large number of users, who might want to only retrieve certain specific data files they are interested in during a given session. One of the most desired ways to do so is through keyword-based search. Such keyword search technique allows users to selectively retrieve files of interest and has been widely applied in plaintext search scenarios [3].

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

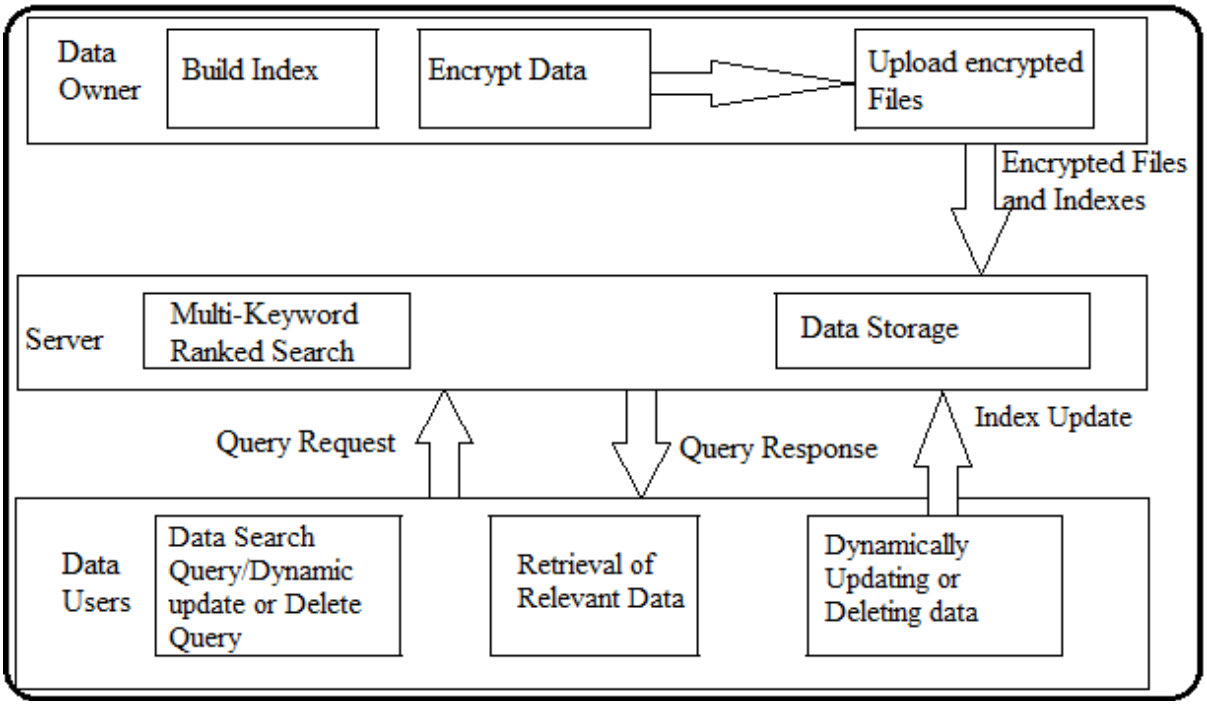

Fig 1: System Architecture

A data owner can outsource their data to the cloud and either he can query on that outsourced data or can authenticate a client to perform query. Various domains where searching is performed on outsourced Cloud data are:

1) **Search Engine**: where a document collection is outsourced to cloud storage and client can retrieve documents which contain the query keywords.

2) **Personalized Medication**: where patient's medical record is outsourced to hospital's server and an authorized doctor can perform secure searching on patient's medical record for diagnosis.

3) **Email Server**: where a collection of private emails is outsourced to email server and client can retrieve pertinent emails based on the content of the mail/sender names/receiver names or email IDs.

4) **Crime Investigation**: where the Interpol's criminal database acts as the server and clients are the authenticated crime investigation agencies like police departments.

The data that is being outsourced may or may not be sensitive. Some of the sentient data might be like patient's medical records, financial data, etc. So, outsourcing plain data will move to some privacy issues. The data owner cannot afford to escape the private data to the CSPs or any unauthorized party. So, for such data owners, their data wants to be encrypted before outsourcing to the third party CSPs. While the encryption of data provides security, it retards the cloud server's ability to perform search operation. Therefore, there is a need for a scheme which can provide a justifiable trade-off between search speed and data privacy. Considering that the large number of data users and documents are available in cloud, it is important for the search service to allow keyword query technique and provide result similarity ranking to meet the practical data retrieval as needed.

MULTI-KEYWORD RANKED SEARCH OVER ENCRYPTED DATA:

To protect data privacy and struggle unwanted 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

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

data use service based on plaintext keyword search. The unimportant 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, excepting 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 powerful 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 huge amount of outsourced data documents in the cloud, this difficulty is mostly demanding as it is really difficult to gather the requirements of performance, system usability, and scalability.

On the other hand, to collect the efficient data retrieval requirement, the huge amount of documents orders the cloud server to reach result relevance ranking, as an alternative of returning undifferentiated results. Such ranked search system permit data users to find the most suitable information quickly, rather than some sorting during every match in the content group. Ranked search can also gracefully remove unwanted 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 release any keyword to related information. To get better the search result accuracy as well as to improve the user searching experience, it is also required for such ranking system to support multiple keywords search, as single keyword search usually give up far too common results. As a regular practice specifies by today's web search engines i,e Google search, data users may slant to offer a set of keywords as an alternative of only one as the indicator of their search interest to retrieve the most applicable data. And each keyword in the search demand is able to help the limited search result further.[5]

II. LITERATURE SURVEY

Zhihua Xia,Xinhui Wang, Xingming Sunand Qian Wang, fellow, IEEE [1]in this paper they work on a secure tree-based search scheme over the encrypted cloud storage, which support multi keyword ranked search across with dynamic operation on document collection available at server. The vector space model and term frequency (TF) × inverse document frequency (IDF) model are combinly used in the creation of index and generation of query to provide multi keyword ranked search output. To get high search efficiency results, author construct a tree-based index structure and proposed a Greedy Depth-first Search algorithm based on this index tree. Because of this special structure of tree-based index, the proposed search scheme can flexibly reach sub linear search time and can effectively deal with the deletion and insertion of documents. The kNN algorithm is applied to encrypt the index and query vectors, and till then make sure accurate relevance score calculation between encrypted index and query vectors. The proposed scheme is designed to provide not only multi-keyword query and correct result ranking, but also dynamic update on document collections. They describe the unencrypted dynamic multi-keyword ranked search (UDMRS) scheme which is construct on the basis of vector space model and KBB tree. Based on the UDMRS scheme, two secure search schemes (BDMRS and EDMRS schemes) are constructed against two threat models, respectively. There are still many challenges in symmetric SE schemes. In the proposed scheme, the data owner is responsible for generating updating information and sending them to the cloud server. Thus, the data owner requires storing the unencrypted index tree and the information that are necessary to recalculate the IDF values. Such an active data owner may not be very suitable for the cloud computing model. It could be an important but difficult future work to design a dynamic searchable encryption scheme whose updating operation can be completed by cloud server only, for now reserving the ability to support multi-keyword ranked search. In addition, as the most of works about searchable encryption, this scheme mainly considers the challenge from the cloud server.

Bing Wang, Shucheng Yu, Wenjing Lou Y. Thomas Hou follow IEEE [2] in this paper they worked on blocked the challenging multi-keyword fuzzy search problem over the encrypted data. They proposed integrated several new designs to solve the multiple keywords search and the fuzzy search problems simultaneously with high efficiency. The approach of holding LSH functions in the Bloom filter to construct the file index is novel and provides an efficient solution to the secure fuzzy search of multiple keywords. In addition, the Euclidean distance is acquire to capture the similarity between the keywords and the secure inner product computation is used to calculate the similarity score so as to enable result ranking. They proposed a basic scheme as well as as better scheme in order to meet different security requirements.

To design a secure and well-functioning search scheme over encrypted data, one has to make three important design choices that are closely inter-related and largely determine the performance of the resulting search scheme, 1). Data

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

structure used to construct secure indexes and trapdoors 2). Effective search algorithm that can quantify the level of match between keywords in the query and keywords in the index with high efficiency, and 3) security and privacy mechanisms that can be mixed in the above two design choices thus the index privacy and search privacy can be protected.

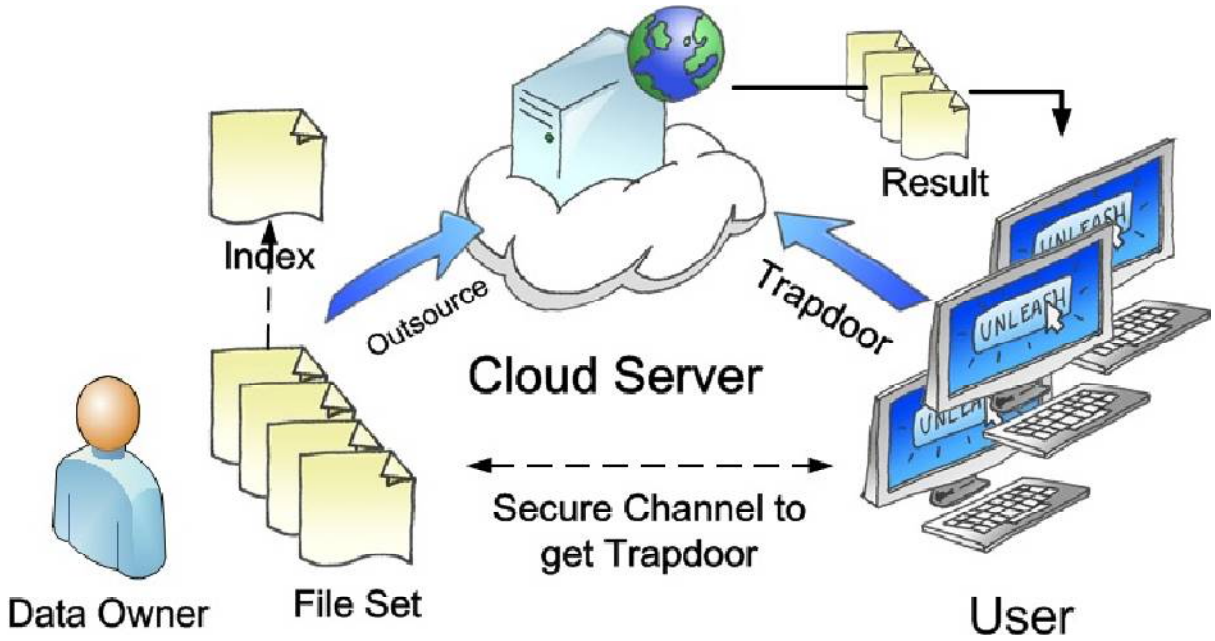

Fig 2: System architecture of search over encrypted data in cloud computing

Cong Wang, Ning Cao, KuiRen, Wenjing Lou, fellow, IEEE [3] in this paper they work on the ranked keyword search over encrypted data to reach economies of scale for Cloud Computing. They start from the study of existing searchable symmetric encryption (SSE) schemes and provide the definitions and framework for their proposed ranked searchable symmetric encryption (RSSE). Note that by following the same security bond of existing SSE, it would be very unsuitableto support ranked search functionality over encrypted data, as demonstrated in our basic scheme. The conversion of its demerits will show the proposed scheme. Searchable encryption allows data owner to outsource his data in an encrypted manner while maintaining the selectively-search ability over the encrypted data. Generally, searchable encryption can be achieved in its full functionality using an insensible RAMs. They first start with aeasy ideal scheme, where the security of their ranked searchable encryption is the same as previous SSE schemes, i.e., the user gets the ranked results without allow cloud server learn any additional information more than the access pattern and search pattern. However, this is achieved with the trade-off of efficiency: namely, either should the user wait for two round-trip time for each search request, or he may even lose the capability to perform top-k retrieval, resulting the unnecessary communication overhead. They believe the analysis of these demerits will lead to their main result. EFFICIENT RANKED SEARCHABLE SYMMETRICENCRYPTION SCHEME proposesthat causes the inefficiency of ranked searchable encryption. That is how to allow server quickly perform the ranking without actually knowing the relevance scores. Note that by resorting to OPSE, our security guarantee of RSSE is inherently weakened compared to SSE, as they now let server know the relevance order. However, this is the information they want to trade off for efficient RSSE, then they show how we can adapt it to suit our purpose for ranked searchable encryption with an “as-strong-as-possible” security warranty. Using Order Preserving Symmetric Encryption [OPSE]deterministic encryption scheme where the numerical ordering of the plaintexts gets preserved by the encryption function.

CengizOrencik, Murat Kantarciogluy and ErkaySavas fellow IEEE [4] in this paper they work on the addressed of privacy-preserving multi-keyword search over encrypted cloud data for the database outsourcing scenario. They present a novel method using min-hash functions that provide logical comparison between signatures of documents and queries. They give formal security definitions and prove that their proposed work satisfies adaptive semantic security.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

They incorporate ranking capacity to the proposed scheme utilizing well known tf-idf based relevancy scoring. This approach ensures that only the most relevant items are retrieved by the user, preventing unnecessary communication and computation load on the user. They implement the entire system and demonstrate the effectiveness and efficiency of their solution through extensive experiments using the publicly available Enron dataset. Most of the secure search methods in literature do not support multiple features in queries. They do not provide any comparison with those single keyword search methods but compare their proposed method with the existing secure multi-keyword search methods instead. Some of the multi-keyword search methods utilize bilinear mapping. This approach has similar security requirements with their proposed method, such that it reveals search and access pattern but nothing else. In this work, each search operation does about 2l bilinear mapping operation where l is the number of features in a document, which is not practical due to the cost of bilinear map operations. A recent work by Cao et al utilizes matrix multiplication operations where the number of rows is determined by the size of the complete feature set. This method performs index construction for 6000 documents in about 4500 s while we perform the same operation in less than 600 s. Similarly, the search operation over 6000 documents requires 600 ms, while we perform in about 210 ms.

III. PROPOSED WORK

Proposed work is based on a secure tree-based search scheme over the encrypted cloud data, which supports multi-keyword ranked search and dynamic operation on the document collection. Specifically, the vector space model and the broadly-used "term frequency (TF) × inverse document frequency (IDF)" model are combined in the index construction and query generation to provide multi-keyword ranked search. The secure KNN algorithm is used to encrypt the index and query vectors, and meanwhile ensure accurate relevance score calculation between encrypted index and query vectors. To oppose different attacks in different threat models, we construct two secure search schemes: the basic dynamic multi-keyword ranked search (BDMRS) scheme in the known cipher text model, and the enhanced dynamic multi-keyword ranked search (EDMRS) scheme in the known background model.

IV. CONCLUSION

Multi-keyword ranked search use a secure, efficient and dynamic search scheme which supports not only the accurate multi-keyword ranked search but also the dynamic deletion and insertion of documents. It construct a special keyword balanced binary tree as the index, and propose a "Greedy Depth-first Search" algorithm to obtain better efficiency than linear search. In addition, the parallel search process can be carried out to further reduce the time cost. The security of the scheme is protected against two threat models by using the secure KNN algorithm. Experimental results demonstrate the efficiency of proposed scheme. There are still many challenge problems in symmetric SE schemes. In that the data owner is responsible for generating updating information and sending them to the cloud server. Thus, the data owner needs to store the unencrypted index tree and the information that are necessary to recalculate the IDF values. Such an active data owner may not be very suitable for the cloud computing model. It could be a meaningful but difficult future work to design a dynamic searchable encryption scheme whose updating operation can be completed by cloud server only, meanwhile reserving the ability to support multi-keyword ranked search. In addition, as the most of works about searchable encryption, our scheme mainly considers the challenge from the cloud server. Actually, there are many secure challenges in a multiuser scheme.

REFERENCES

1) Zhihua Xia, Member, Xinhui Wang, Xingming Sun and Qian Wang, Member, IEEE, "A Secure and Dynamic Multi-keyword Ranked Search Scheme over Encrypted Cloud Data", IEEE transactions on Parallel and Distributed systems,2015
2) B. Wang, S. Yu, W. Lou, and Y. T. Hou, "Privacy-preserving multi-keyword fuzzy search over encrypted data in the cloud," in IEEEINFOCOM, 2014.
3) C. Wang, N. Cao, K. Ren, and W. Lou, "Enabling secure and efficient ranked keyword search over outsourced cloud data," IEEE Transactions on Parallel and Distributed Systems, vol. 23,2015.
4) C. Orencik, M. Kantarcioglu, and E. Savas, "A practical and secure multi-keyword search method over encrypted cloud data," in Cloud Computing (CLOUD), 2013 IEEE Sixth International Conference on. IEEE, 2013.
5) C. Gentry, "A fully homomorphic encryption scheme," Ph.D. dissertation, Stanford University, 2009.
6) Y.-C. Chang and M. Mitzenmacher, "Privacy preserving keyword searches on remote encrypted data," in Proceedings of the Third internationalconference on Applied Cryptography and Network Security. Springer-Verlag, 2005.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

7) M. Kuzu, M. S. Islam, and M. Kantarcioglu, "Efficient similarity search over encrypted data," in Data Engineering (ICDE), 2012IEEE 28th International Conference on. IEEE, 2012.
8) 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 Systems and Networks (DSN), 2014 44th AnnualIEEE/IFIP International Conference on. IEEE, 2014.
9) C. Wang, K. Ren, S. Yu, and K. M. R. Urs, "Achieving usable and privacy-assured similarity search over outsourced cloud data," in INFOCOM, 2012 Proceedings IEEE. 2012,