SECURE ATTRIBUTE- BASED SEARCH SCHEMA FOR DATA STORED IN CLOUD

Murthy Sowmya M.Tech(CSE),
Kumar Devapogu Associate Professor & Dept.of CSE
Dr.Samuel George Institute of Engineering and Technology,Markapur,India

Abstract

Now-a-days, most of the users store their personal and professional data on the cloud. Due to the massive increase in the storage and computing requirements of users, every time, data is getting transferred to the remote server in larger chunks, without an analysis whether the server on which the data is outsourced, is a trusted server. But the fact is, after outsourcing, users are with a great security risk factor that tends to lose the local possession of their large size of data. To maintain the privacy of personal documents stored in cloud environment, it should get encrypted before outsourcing to the cloud. After placing the data on the cloud, retrieving the same is also a quiet tedious job. In order to retrieve the data, several approaches are available in which keyword enabled search of the encrypted data is one of the outstanding techniques. The majority of these approaches are limited to handle a single keyword search with its own limitation. To enhance searching in terms of efficiency and fastness, a multi-key word search technique can be adopted to retrieve a corresponding document from cloud. This paper proposes a survey on a secure search scheme supporting single-keyword or multi-keyword ranked search over encrypted cloud data.

Keywords— Single keyword, Multi-keyword search, Ranked search, Encrypted cloud data, Security

1.INTRODUCTION

Cloud Computing is an increasing mature model of enterprise IT infrastructure that provides on demand high quality applications and services from a shared pool of configuration computing resources. The cloud customers, individuals or enterprises, can outsource their local complex data system into the cloud to avoid the costs of building and maintaining a private storage infrastructure. The company or organization's private and sensitive information like personal files, company records, emails, etc which is to be shared among the selected company employees is stored and centralized into cloud server but mostly with an insecure feeling that anyone may hack these data that may be very risky for that company. Also the data owners and cloud server may not to be in the same trusted domain who put the outsourced unencrypted data, if any, at risk; the cloud server may leak data information to unauthorized entities or even be hacked.

Cloud enables large group of remote servers to be in a network so as to allow the centralized data repository, and access to the computer services or resources whenever required. Many users are motivated to outsource their confidential data on to the cloud. As the documents get transferred to the cloud, users don't have physical possession of that data. So as to make sure that the data at cloud side is safer, it has to adapt to the privacy preserving storage, as the cloud server is not a trusted server. To protect data confidentiality and unauthorized access to the cloud data, owners are motivated to encrypt their data before it is outsourced to cloud. To overcome this issue, the data stored in cloud storage database needs to be encrypted before sending to cloud servers for storage.

Due to the significance of effective storage and computational models available in the cloud, number of cloud service users is increasing day by day. As a result of this, bulks of information are being getting added into the cloud servers. In such environment, searching and retrieving the required file is like finding needle onto the sand beach. Sometimes, due to encrypted forms of data on cloud may also force us to have to search through it, in encrypted form only. This makes the retrieving problem a challenging task. Retrieving files without relevance makes wastage of

computational cost and unreliable way to access files. Thus, the use of efficient search technique gains a very high importance nowadays. As user stores the encrypted data at cloud side, traditional searching will not be effective as well. To meet the effective searching on the encrypted cloud data, multi-keyword query can be formed so as to get the top relevant data of user interest.

Searching Techniques:
There are various searching techniques available, and to mention a few are as follows:

- **Searchable Encryption:** It allows users to securely search complete encrypted data through keywords. This method support only Boolean search, without capturing any relevant data. This approach suffers from two main drawbacks when directly applied in the context of Cloud Computing. First one, users who do not necessarily have pre-knowledge of the encrypted cloud data, have to post process every file got, in order, to find ones most matching their interest; another drawback, regularly getting all files containing the queried keyword further incurs unnecessary network traffic, when retrieve more than one files.

- **Single Keyword Searchable Encryption:** A single keyword searchable encryption schemes usually builds an encrypted searchable index such that, it's content is hidden to the server, unless it is given appropriate trapdoors generated via secret key(s). Early work solves secure ranked keyword search which utilizes keyword frequency to rank results instead of returning undifferentiated results. However, it only supports single keyword search. Where anyone with public key can write to the data stored on server, but only authorized users with private key can search. Traditional single keyword searchable encryption schemes are usually built in a way by creating an encrypted searchable index. Such indexes content will be hidden to the server. The information will be revealed only when the server gives the correct trapdoors that are generated via a secret key(s). The main drawback of single keyword-based search is that it is not comfortable enough to express complex information needs.

- **Ranked Keyword Search:** Ranked search greatly enhances system usability by returning the matching files in a ranked order regarding to certain relevance criteria (eg. keyword frequency) thus, making one step closer toward practical deployment of privacy-preserving data hosting services in the context of cloud computing. To the best of knowledge it gives a legal status for the first time the problem of effective ranked keyword search over encrypted cloud data. Ranked keyword search strongly provides system usability by returning the matching files in ranked order concerning to certain relevance criteria, thus moving close towards the practical action of privacy preserving data presenting services in cloud computing. To achieve design goals investigate the statistical measure approach from Information retrieval (IR) and text removal to insert relevance score of each file during the establishment of searchable index before outsourcing the encrypted file collection. An IR system allocates a relevance score to each and every document and ranks those documents by this score. Relevance score is used to build a secure searchable index to properly protect the sensitive information. This technique enables data users to find the most related information rapidly, rather than burdensome sorting through every match in the content collection. Ranked search can also elegantly eliminate unnecessary network traffic by sending back only the most relevant data. For privacy protection, such ranking function, however, should not leak any keyword relevant information. Another one, to improve search result accuracy as well as enhance user searching experience, it is also essential for such ranking system to support multiple keywords search.

- **Multi-keyword Searchable encryption:** Over the years, various searchable encryption approaches have been developed toprovide the ability for selectively retrieving the encrypted documents through

a keyword search. Typically, these systems build a secure index structure and outsource it along with the encrypted documents to the remote server. Authorized users submit their requests as secret trapdoors that are integrated properly with the stored indexing information. The server uses the received trapdoor to search over the stored index, and retrieves the matching encrypted documents. However, the previous searchable encryption schemes are impractical for real world cloud computing scenarios because these systems are designed to handle either a single keyword search or a Boolean search.

- **Fuzzy Keyword Searchable Encryption:** Fuzzy keyword search greatly enhances system usability by returning the matching files when users searching inputs exactly match the predefined keywords or the closest possible matching files based on keyword similarity semantics, when exact match fails. More specifically, it uses edit distance to quantify keywords similarity and develop a novel technique that is a wildcard-based technique, for the construction of fuzzy keyword sets. This technique eliminates the need for enumerating all the fuzzy keywords and the resulted size of the fuzzy keyword sets is significantly reduced.

- **Plaintext Fuzzy Keyword Search:** The importance of fuzzy search has received attention in the context of plaintext searching in information retrieval community. The problem is addressed in the traditional information access paradigm by allowing user to search without using try-and-see approach for finding relevant information based on approximate string matching. At the first glance, it seems possible for one to directly apply these string matching algorithms to the context of searchable encryption by computing the trapdoors on a character base within an alphabet. This trivial construction suffers from the dictionary and statistics attacks and fails to achieve the search privacy.

- **Boolean Keyword Search:** Boolean systems allowed customers to specify their information need using a combination of Boolean operators AND, OR and NOT. Boolean systems have several disadvantages, for example there are no any features of document ranking, and it is very difficult for a customer to make a good search request. Thus, the drawback of existing system specifies the important need for new techniques that support searching flexibility.

2. LITERATURE SURVEY

Even though there are various systems existing, this literature survey mainly concentrates on the single keyword based encryption and multi-keyword based encryption and also included other searching techniques due to it known advantages.

1) Single keyword search:

- Deepali D. Rane et.al, [1] proposed implementation of the encryption and decryption, Secure index construction is successfully completed with desirable performance. After index construction it will get compressed and will be stored in
.cfs file format. After firing single-keyword query, user will get all documents that contain the specified keyword. The advantages are protects data privacy by encrypting documents before outsourcing, rank based retrieval of the documents, To easily access the encrypted data by multi keyword rank search using keyword index. The Disadvantages of the proposed system are single-keyword search without ranking, Boolean keyword searching without ranking, single-keyword search with ranking, Rarely sorting of the results i.e. no index creation and ranking, Single User search.

- C. Wang et al, [2] proposed a secure ranked keyword search technique that utilizes the keyword frequency to rank the results. The major drawbacks of single keyword search systems with or without ranking most probably won't retrieve the relevant data and it may also compromise the privacy.

- Y.-C. Chang et al, [3] proposed similar "index" approaches, in which a single encrypted hash table index was built for the entire file collection.

- D. Song, D. Wagner et al, [4] proposed a searchable encryption, in which each and every word in the considered document was encrypted autonomously under two-layered encryption construction.

3.Multi-keyword search:

- Zhihua Xia et.al,[5] proposed 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. They construct a special keyword balanced binary tree as the index, and proposed 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. The Advantages of the proposed system are searchable encryption schemes enable the client to store the encrypted data to the cloud and execute keyword search over cipher text domain and a secure tree-based search scheme over the encrypted cloud data, which supports multi-keyword ranked search and dynamic operation on the document collection. The disadvantages are 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.

- Bing Wang et.al, [6] proposed a novel construction of a public key searchable encryption scheme based on inverted index. This scheme overcomes the one-time-only search limitation in the previous schemes. The disadvantages of the proposed system are first of all, the keyword privacy is compromised once a keyword is searched. As a result, the index must be rebuilt for the keyword once it has been searched. Such solution is counterproductive due to the high overhead suffered. Secondly, the existing inverted index based searchable schemes do not support conjunctive multi-keyword search, which is the most common form of queries now a days. The advantages are explore the problem of building a searchable encryption scheme based on the inverted index, Achieve secure and private matching between the query trapdoor and the secure index, Design a novel trapdoor generation algorithm so that the query related inverted lists are combined together secretly without letting the cloud server know which inverted lists are retrieved.

- Yanzhi Ren et.al, [7] proposed a light-weight search approach that supports efficient multi-keyword ranked search in cloud computing system. The basic scheme employs the polynomial function to hide the encrypted keyword and search patterns for efficient multi-keyword ranked search. Then improve the basic scheme and propose a privacy-preserving scheme which utilizes the secure inner product method for protecting the privacy of the searched multi-keywords. The advantage of the proposed system is it analyzes the privacy guarantee of the proposed scheme and conduct extensive experiments based on the real-world dataset. The disadvantage is there is a chance of leakage of data in cloud.

- Hongwei Li et.al, [8] proposed a multi-keyword ranked search scheme to enable accurate, efficient and secure search over encrypted mobile cloud data. Security analysis have demonstrated that proposed scheme can effectively achieve confidentiality of documents and index, trapdoor privacy, trapdoor unlinkability, and concealing access pattern of the search user. The advantages Constructs an efficient index to improve the search efficiency. And it solves the trapdoor unlinkability problem. It also achieve enhanced efficiency in terms of functionality and search efficiency compared with existing proposals.

- Mikhail Strizhov et.al, [9] proposed a searchable encryption technique that enables secure searches over encrypted data stored on remote servers. They define and solve the problem of multi-keyword ranked search over encrypted cloud data. In particular, they present an efficient similarity searchable encryption scheme that supports multi-keyword semantics. The solution is based on two building blocks: Term Frequency Inverse Document Frequency (TF-IDF) measurement and ring-LWE-based variant of homomorphic cryptosystem. The Advantages of this system is it returns the matching data items in a ranked ordered manner. The Disadvantage is in traditional system it supports only single keyword search.

4. Other searching techniques:

- E.-J. Goh et al, [10] proposed a technique that uses Bloom filters in order to construct the indexes for the data files. Bloom filter containing trapdoors (for each file) of all distinct words is built up and stored on the server. For searching a particular word, the user must generate the search request by computing the trapdoor of the word and sends it to the server. The server upon receiving the request performs tests to check if any Bloom filter holds the trapdoor of the query word and if so, it returns the corresponding file identifiers.

- Jun Zhou et.al, [11] proposed a more efficient verifiable outsourced computation of encrypted data EVOC from any one- way trapdoor function is proposed by combining a newly devised privacy-preserving data aggregation supporting both addition and multiplication operations with Yao's Garbled Circuit. The advantage is it proves the security of the proposed efficient privacy-preserving data aggregation scheme.

- Fanyu Bu et.al, [12] proposed a privacy preserving backpropagation algorithm based on the BGV encryption scheme on cloud. One property of the proposed algorithm is to apply the BGV encryption scheme to the back-propagation algorithm for preventing the disclose of private data with cloud computing. The advantages: The proposed algorithm improves the efficiency of back-propagation learning by offloading the expensive operations on the cloud. It also prevents the disclosure of private data, using full homomorphic encryption scheme to encrypt the source data. The disadvantage is sensitive data iseasily disclosed during the process of the computation on the cloud.

- Joseph K et.al, [13] proposed an infrastructure for secure sharing and searching for real-time video data. It is particularly suitable for mobile users by deploying 5G technology and a cloud computing platform. The security is guaranteed even if the cloud server is hacked since data confidentiality is now protected by cryptographic encryption algorithms. The advantage of the proposed system is the infrastructure security is guaranteed even if the cloud server is hacked. The disadvantage is There are some existing platforms for sharing real time video, they may not be able to achieve secure fine-grained sharing and secure searching simultaneously.

- Zhangjie Fu et.al, [14] proposed an efficient verifiable keyword-based semantic search scheme. Comparing to most of the existing searchable encryption schemes, the proposed scheme is more practical and flexible, better suiting user's different search intentions. Moreover, the proposed scheme protects data privacy and supports verifiable search ability, in the presence of the semihonest server in the cloud computing environment. The Advantages: Improves the flexibility and support verification of search results. Also it provides verifiable searchability with data privacy preserving. The disadvantage is The trivial solution of downloading the whole encrypted data first and then decrypting it locally is obviously impractical, due to the huge bandwidth and computation burden.

- Jin Li et al, [15] proposed a new searching technique that is fuzzy keyword search. They focused on enabling effective yet privacy preserving fuzzy keyword search in Cloud Computing. To the best of knowledge, they formalize for the first time the problem of effective fuzzy keyword search over encrypted cloud data while maintaining keyword privacy.

- D. Boneh, G. D. Crescenzo et.al, [16] proposed a searchable encryption, in which anyone who has the public key can write to the data stored in server but with a restriction that only authorized users with private key can search. The major demerits of using public key are that it's very expensive in terms of computation. Additionally, in the public key setting, the privacy of keyword may possibly not be protected as the server possesses the ability to encrypt any keyword with the public key. As a result of this it can be used to receive the trapdoor in order to evaluate the cipher text. For achieving more efficient search, Curtmola et al. single keyword search.

5. CONCLUSION

This paper proposes brief literature survey on single keyword based searching, multi-keyword based searching and many other searching techniques in cloud based environment. From the survey, the multi-keyword search technique sounds to be more efficient than other available searching technique. Many search schemes over encrypted data, supports multi-keyword query and similarity ranking simultaneously for data retrieval in cloud computing.

6. REFERENCES

[1] Deepali D. Rane and Dr.V.R.Ghorpade " Multi-User Multi-Keyword Privacy Preserving Ranked Based Search Over Encrypted Cloud Data" International Conference on Pervasive Computing (ICPC), 2015.

[2] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure ranked keyword search over encrypted cloud data," in Proc. of ICDCS'10,2010.

[3] Y.-C. Chang and M. Mitzenmacher, "Privacy preserving keyword searches on remote encrypted data," in Proc. of ACNS, 2005.

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

[5] Zhihua Xia, Member, IEEE, Xinhui Wang, Xingming Sun, Senior Member, IEEE, 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, VOL., NO.1,2015.

[6[Bing Wang, Wei Song, Wenjing Lou, and Y. Thomas Hou "Inverted Index Based Multi-Keyword Public-key SearchableEncryption with Strong Privacy Guarantee" IEEE Conference on Computer Communications (INFOCOM), 2015. [7]Yanzhi Ren, Yingying Chen, Jie Yang, Bin Xie "

Privacy-preserving Ranked Multi-Keyword Search Leveraging PolynomialFunction in Cloud Computing" Globecom Communication and Information System Security Symposium 2014.

[8] Hongwei Li, Dongxiao Liu,Yuanshun Dai, Tom H. Luan, And Xuemin (Sherman) Shen "Enabling Efficient Multi-Keyword Ranked Search Over Encrypted Mobile Cloud Data Through Blind Storage", December 2014.

[9] Mikhail Strizhov and Indrajit Ray "Multi-keyword Similarity Search Over Encrypted Cloud Data" International Conference on Pervasive Computing (ICPC), 2012.

[10] E.-J. Goh, " Bloom filters in order to construct the indexes for the data files" IEEE Conference on Computer Communications 2016.

[11] Jun Zhou, Zhenfu Cao, Xiaolei Dong and Xiaodong Lin "More Efficient Verifiable Outsourced Computation from Any One- way Trapdoor Function" IEEE ICC - Communication and Information Systems Security Symposium, 2015.

[12] Fanyu Bu, Yu Ma, Zhikui Chen and Han Xu "Privacy Preserving Back-Propagation Based on BGV on Cloud" 2015 IEEE 17th International Conference on High Performance Computing and Communications (HPCC), 2015 IEEE 7th International Symposium on Cyberspace Safety and Security (CSS), and 2015 IEEE 12th International Conf on Embedded Software and Systems (ICESS).

[13] Joseph K, "Secure Sharing and Searching for Real-Time Video Data in Mobile Cloud" 2015.

[14] Zhangjie Fu, *Member,* IEEE, Jiangang Shu, Xingming Sun, and Nigel Linge "Verifiable Keyword-based Semantic Search over Encrypted Cloud Data" IEEE Transactions on Consumer Electronics, Vol. 60, No. 4, November 2014.

[15] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, "Fuzzy keyword search over encrypted data in cloud computing," in Proc. of IEEE INFOCOM'10 Mini-Conference, San Diego, CA, USA, March 2010.

[16] D. Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in Proc. of EUROCRYPT, IEEE Conference on Computer Communications 2004.