A Survey on Multi-Keyword Ranked Search over Encrypted Cloud Data with Multile Data Owners

Kalyani Sonawane
Department of Computer Engineering
Marathwada Mitra Mandal's College of Engineering, Pune.

Rahul Dagade
Professor
Department of Computer Engineering
Marathwada Mitra Mandal's College of Engineering, Pune.

ABSTRACT
The promising benefit of cloud computing is outsourcing of data service, by which the data owners stores their data in the public data centers by economically saving their capital investment towards data management. Cloud Storage provides users with abundant storage space and makes it user friendly for immediate acquiring of data, which is the foundation of all kinds of cloud applications. Data outsourcing in the commercial public cloud also raise the problem for unauthorized data access and the cloud storage does not make sense if the outsourced data is not effectively utilized. The practical challenge is on how to make effective data access in the public cloud storage aiming at improvement of various searching techniques for increasing the data utilization. In this paper, an attempt is made to survey various searching techniques towards effective data utilization in cloud storage and is discussed in detail.

Keywords
Cloud computing, data outsourcing, cloud storage, data utilization.

1. INTRODUCTION
Cloud provides 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 IT enterprises and individuals are outsourcing their databases to cloud server. Variety of users can access and share information stored in the cloud independent of locations. The outsourced data may contain very sensitive information such as e-mails, company financial data, government documents, Personal Health Care records, facebook photos and business documents.

Cloud service providers (CSPs) can access user's sensitive data without any authorization. General approach of CSPs is to protect the data confidentiality in which data is encrypting before outsourcing it to cloud server and this will affect a huge cost of data usability. In secure search over encrypted data, data owners outsourced their data to cloud server in encrypted form to preserve their privacy. When data user wants to search any file, data user send keyword request to cloud server. Cloud server then generate top relevant results to data user. Secure search over encrypted data not only reduce computation cost and storage cost for secure keyword search but also support multi-keyword ranked search, fuzzy keyword search and similarity search. All these schemes are limited to single-owner model.

Earlier work support single-owner model, where data owner has to stay online to generate trapdoors for data user. Therefore, this paper proposes a multi-owner model to overcome the limitations of the earlier methods, where encrypted data are stored by multiple data owners and simultaneously data owners stay online to generate trapdoors. Different data owners share different secret keys to encrypt their secret data with different secret keys.

In this paper, secure search protocol is propose in which cloud server can perform secure search without knowing the actual value of keywords and trapdoors. In multi-owner and multi-user cloud computing model, four entities are involved such as data owners, data users, cloud server and administration server shown in fig 1. Data owners have collection of files. Data owners build secure searchable index of keyword set and keywords are extracted from files. Data owners submit keyword index to administration server. Data owners encrypt files and outsource encrypted files to cloud server. When administration server receives encrypted keyword index then administration server re-encrypt keyword index. Administration server then outsource re-encrypted keyword index to the cloud server.

When data user wants to search over files from cloud server, he first computes the corresponding trapdoors and submits them to the administration server. Administration server authenticates data user then re-encrypts trapdoors and submit them to cloud server. Cloud server searches encrypted index of data owner and returns top-k relevant encrypted files to the data user. When data user receives top-K files from cloud server, then data user download files and decrypts these files.

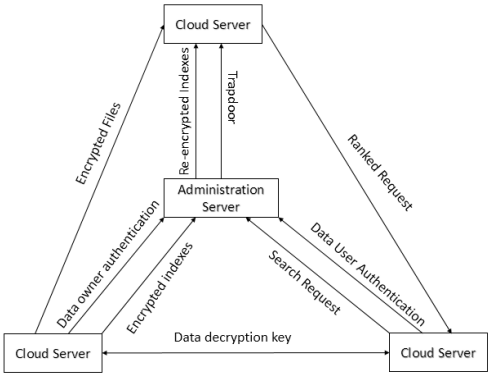

Figure 1 Multi-keyword ranked search over encrypted data

2. LITERATURE SURVEY
In recent years, many researchers have proposed large number of efficient searching schemes over encrypted cloud data. The general process of search scheme is divided into five steps: extracting document features, constructing a searchable index, generating search trapdoor, searching the index based on the trapdoor and returning the search results. These search schemes provide different query capabilities, including single

keyword search, multi-keyword search, fuzzy keyword search, similarity search, and so on.

Secure search over encrypted data have been previously applied to cloud server. Wang et al. [19] proposed secure search scheme over encrypted cloud data. In searchable encryption, clients store data into encrypted form to the cloud server and keyword searching can be perform on ciphertext. Searchable encryption (SE) techniques [4], [5], [6], [7], [13], [14], [15], [16], [17] can partially fulfill the need for secure outsourced data search. Secure search over encrypted cloud data reduces the computation and storage cost. Data user authentication technique, Different-key encrypted keywords matching and privacy preserving ranked search of files methods are used to solve the problem of secure multi-keyword search for multiple data owners and multiple data users in cloud computing environment.

Early works mostly only support single keyword search [16]. Later, several multi-keyword search schemes were proposed [4], [5], [6], [7], [11], [12], [14]. When huge amount of data owners [2], [8] are involved then they generate trapdoors simultaneously which affect the flexibility and usability of search system. Data owner store data in encrypted form and data user generate trapdoors [2], [3], [18] to send query request in encrypted form. Re-encryption of keyword index and trapdoors [8], [10] used to increase more security from attackers.

Zhihua Xia [10] proposed a scheme which supports dynamic update operations like deletion of documents and insertion of documents and tree-based index structure and Greedy Depth first Search algorithm use to provide efficient multi-keyword ranked search. Hongwei Li [11] support complicated logic search by using the mixed AND, OR and NO operations of keywords for practical and very efficient multi-keyword search scheme. [20] Proposed problem of personalized multi-keyword ranked search over encrypted cloud data. A user interest model is build for individual user with the help of semantic ontology WordNet by using user search history.

In searchable symmetric encryption schemes, due to large number of documents, search results should be retrieved in an order of the relevancy with the searched keywords. Scoring is the natural way to weight the relevancy of the documents. TFI-DF [4], [5], [6], [7], [18] is well-known method to compute the relevance score. Wong et al. [12] proposed a secure k-nearest neighbor (kNN) scheme which can confidentially encrypt two vectors and compute Euclidean distance of them [5], [7], [10], [11], [19].

Table 1 Comparative Study

Sr. No.	Paper/Publication	Authors	Methods
1	"A Secure and Dynamic Multi-Keyword Ranked Search Scheme over Encrypted Cloud Data", IEEE Transactions on Parallel and Distributed Systems, February 2016 [10]	Zhihua Xia, Xinhui Wang, Xingming Sun, Qian Wang	Scheme supports dynamic update operations like deletion of documents and insertion of documents. Tree-based index structure and "Greedy Depth First Search" algorithms are use to provide efficient multi-keyword ranked search.
2	"Enabling Fine-Grained Multi-Keyword Search Supporting Classified Sub-Dictionaries over Encrypted Cloud Data", IEEE Transactions on Dependable and Secure Computing, May/June 2016 [11]	Hongwei Li, Yi Yang, Tom H. Luan, Xiaohui Liang, Liang Zhou	Relevance scores and preference factors of keywords use to enable precise keyword search and personalized user experience. Support complicated logic search by using the mixed "AND", "OR" and "NO" operations of keywords. Classified sub-dictionaries technique is used to achieve better efficiency on index building, trapdoor generating and query.
3	"Enabling Personalized Search over Encrypted Outsourced Data with Efficiency Improvement", IEEE Transactions on Parallel and Distributed Systems, September 2016 [20]	Zhangjie Fu, Kui Ren, Jiangang Shu, Xingming Sun, Fengxiao Huang	By using the user search history, a user interest model is build for individual user with the help of semantic ontology WordNet. The user interest model is use to realize automatic evaluation of the keyword priority and it solved the limitation of the artificial method of measuring.
4	"An Efficient File Hierarchy Attribute Based Encryption Scheme in Cloud Computing", IEEE Transactions on Information Forensics and Security, June 2016 [22]	Shulan Wang, Junwei Zhou, Joseph K. Liu, Jianping Yu, Jianyong Chen	Uses Ciphertext-policy attribute-based encryption (CP-ABE) encryption technology to solve the challenging problem of secure data sharing in cloud computing. Efficient file hierarchy attribute-based encryption scheme is proposed in cloud computing.
5	"Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data", IEEE Transactions on Information Forensics and Security, January 2014 [5]	Shulan Wang, Junwei Zhou, Joseph K. Liu, Jianping Yu, Jianyong Chen	Propose two MRSE schemes based on the similarity measure of "coordinate matching" to provide as many matches as possible to effectively capture the relevance of outsourced documents to the query keywords while meeting different privacy requirements. "Inner product similarity" is used to quantitatively evaluate similarity measure.

3. DIFFERENT TECHNIQUES TO SERACH OVER ENCRYPTED CLOUD DATA

3.1. Search over Encrypted Data With Authorization Framework:

The search authorization framework adds another layer of fine-grained privacy protection for data access control over encrypted cloud data. [2] Data owners and data users do not directly interact with each other. Trusted Authority (TA) and Local Trusted Authorities (LTAs) provide privileges to cloud users. [3], [13] TPA handle multiple audit session from different users also perform multiple auditing tasks in a batch manner for better efficiency.

3.2. Secure Index

The secure index scheme builds a secure index for keywords extracted from documents. This secure index allows a user to search for an encrypted document that is containing a keyword without decrypting the document. [4] Tree based index structure used to store keywords so that search efficiency is much better than linear search. [10] Propose a "Greedy Depth First Search" algorithm to provide efficient search over special tree based index structure. Inverted index [12] is most efficient searchable index structure and mostly support to plaintext search.

3.3. Similarity Search over Encrypted Data:

Documents are encrypted before stored to cloud server so authorized users are allowed to access cloud data. There are different searching techniques are available which handle only exact query matching. [4], [5], [6] not only handles exact query matching but also matches query based on its similarity with documents. Documents are retrieved if its similarity against a specified query word is greater than or equal to predefined threshold.

3.4. Public Key Encryption with Keyword Search

Public key encryption [12] is encryption scheme in which cloud server contains encrypted files and keyword index. Users create trapdoors by using its private key. The cloud server checks the trapdoor with existing encrypted keyword and sends back encrypted files that match it.

3.5. Practical Techniques for Searches on Encrypted Data

The scheme is based on sequential scan method. PTSED consists of several steps: Pre-encryption, searching, and decryption. The purpose of the pre-encryption first step is to hide the actual searching keyword and to prevent any unauthorized party which can excess the remote server using cryptanalysis to break the whole encrypted message after a few keyword searches. Before starting the searching algorithm, the user has to provide some information since the server will not learn anything more than what is provided by the user. After the server gathers the required information from the user, the searching algorithm will run based on the information gathered. In this case, the server may return the file to the end user if the keyword is match. Otherwise, it will continue to search until the end of the file. After the user search and retrieve the encrypted file containing the specific keyword, the final step is to decrypt the retrieved file back to plaintext [21].

3.6. Multi-keyword Search over Encrypted Data with Multiple Data Owners

Most cloud servers just serve one data owner. First, in the single-owner scheme, the data owner has to stay online to generate trapdoors for data users. When a huge amount of data owners are involved, asking them to stay online simultaneously to generate trapdoors would seriously affect the flexibility and usability of the search system. Second, none of us would be willing to share our secret keys with others, different data owners would prefer to use their own secret keys to encrypt their secret data [1], [2].

4. CONCLUSION

This paper summarizes various searching techniques in the encrypted cloud data. 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 information behind the data collection. Performing such kind of searching causes an increase in the computational cost and the cost associated with communication. 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.

5. REFERENCES

[1] Wei Zhang, Yaping Lin, Sheng Xiao, Jie Wu, Siwang Zhou, "Privacy Preserving Ranked Multi-Keyword Search for Multiple Data Owners in Cloud Computing", IEEE Transactions on Computers, Vol. 65, No. 5, May 2016.

[2] Ming Li, Shucheng Yu, Ning Cao, Wenjing Lou, "Authorized Private Keyword Search over Encrypted Data in Cloud Computing", 31st International Conference on Distributed Computing Systems, 2011.

[3] Cong Wang, Sherman S.M. Chow, Qian Wang, Kui Ren, Wenjing Lou, "Privacy-Preserving Public Auditing for Secure Cloud Storage", IEEE Transactions on Computers, Vol. 62, No. 2, February 2013.

[4] Wenhai Sun, Bing Wang, Ning Cao, Ming Li, Wenjing Lou, Y. Thomas Hou, Hui Li, "Verifiable Privacy-Preserving Multi-Keyword Text Search in the Cloud Supporting Similarity-Based Ranking", IEEE Transactions on Parallel and Distributed Systems, Vol. 25, No. 11, November 2014.

[5] Ning Cao, Cong Wang, Ming Li, Kui Ren, Wenjing Lou, "Privacy Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data", IEEE Transactions on Parallel and Distributed Systems, Vol. 25, No. 1, January 2014.

[6] Zhangjie Fu, Xingming Sun, Zhihua Xia, Lu Zhou, Jiangang Shu, "Multikeyword Ranked Search Supporting Synonym Query over Encrypted Data in Cloud Computing", 2013.

[7] Zhangjie Fu, Xingming Sun, Nigel Linge, Lu Zhou, "Achieving Effective Cloud Search Services: Multi-keyword Ranked Search over Encrypted Cloud Data Supporting Synonym Query", IEEE Transactions on Consumer Electronics, Vol. 60, No. 1, February 2014.

[8] Wenhai Sun, Shucheng Yu, Wenjing Lou, Y. Thomas Hou, Hui Li, "Protecting Your Right: Verifiable Attribute-Based Keyword Search with Fine-Grained Owner-Enforced Search Authorization in the Cloud", IEEE Transactions on Parallel and Distributed Systems, Vol. 27, No. 4, April 2016.

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

[10] Zhihua Xia, Xinhui Wang, Xingming Sun, Qian Wang, "A Secure and Dynamic Multi-Keyword Ranked Search Scheme over Encrypted Cloud Data", IEEE Transactions on Parallel and Distributed Systems, Vol. 27, No. 2, February 2016.

[11] Hongwei Li, Yi Yang, Tom H. Luan, Xiaohui Liang, Liang Zhou, Xuemin (Sherman) Shen, "Enabling Fine-Grained Multi-Keyword Search Supporting Classified Sub-Dictionaries over Encrypted Cloud Data", IEEE Transactions on Dependable and Secure Computing, Vol. 13, No. 3, May/June 2016.

[12] Bing Wang, Wei Song, Wenjing Lou, Y. Thomas Hou, "Inverted Index Based Multi-Keyword Public-key Searchable Encryption with Strong Privacy Guarantee", 2015 IEEE Conference on Computer Communications(INFOCOM).

[13] Wenhai Sun, Xuefeng Liu, Wenjing Lou, Y. Thomas Hou, Hui Li, "Catch You If You Lie to Me: Efficient Verifiable Conjunctive Keyword Search over Large Dynamic Encrypted Cloud Data", 2015 IEEE Conference on Computer Communications (INFOCOM).

[14] Hongwei Li, Dongxiao Liu, Kun Jia, Xiaodong Lin, "Achieving Authorized and Ranked Multi-keyword Search over Encrypted Cloud Data", IEEE ICC 2015-Communication and Information Systems Security Symposium.

[15] Wei Zhang, Yaping Lin, "Catch You if You Misbehave: Ranked Keyword Search Results Verification in Cloud Computing", VOL. 6, NO. 1, JANUARY 2015.

[16] R. Curtmola, J. A. Garay, S. Kamara, R. Ostrovsky, "Searchable symmetric encryption: improved definitions and efficient constructions", In Proc. of ACM CCS 06, 2006.

[17] C. Wang, N. Cao, J. Li, K. Ren, W. Lou, "Secure ranked keyword search over encrypted cloud data", in Proc. IEEE Distributed Computer System, Genoa, Italy, Jun. 2010, pp. 253262.

[18] Hongwei Li, Dongxiao Liu, Yuanshun Da11i, Tom H. Luan, Xuemin Shen, "Enabling Efficient Multi-Keyword Ranked Search Over Encrypted Mobile Cloud Data Through Blind Storage", Transaction On Emerging Topics In Computing, 6 March, 2015.

[19] W. K. Wong, D. W. Cheung, B. Kao, and N. Mamoulis, "Secure kNN computation on encrypted databases", in Proc. ACM SIGMOD Int. Conf. Manage. Data, 2009, pp. 139-152.

[20] Zhangjie Fu, Kui Ren, Jiangang Shu, Xingming Sun, Fengxiao Huang, "Enabling Personalized Search Over Encrypted Outsourced Data With Efficiency Improvement", IEEE Transactions on Parallel and Distributed Systems, Vol. 27, No. 9, September 2016.

[21] Dawn Xiaodong Song David Wagner Adrian Perrig, "Practical Techniques for Searches on Encrypted Data", University of California, Berkeley.

[22] Shulan Wang, Junwei Zhou, Joseph K. Liu, Jianping Yu, Jianyong Chen, Weixin Xie, "An Efficient File Hierarchy Attribute-Based Encryption Scheme in Cloud Computing", IEEE Transactions on Information Forensics and Security, Vol. 11, No. 6, June 2016.