ISSN 2395-1621

Literature Survey on Multi-keyword Ranked Search for Multiple Data Owners in Cloud Computing

[#1]Mrs. D.A. Phalke, [#2]Sayali Sabale

[1]a_dhanashree@rediffmail.com
[2]sayala.sabale@gmail.com

[#12]Department of Computer Engineering

D.Y. Patil College of Engineering, Akurdi, Pune - 411044

ABSTRACT

Abstract- In cloud computing, a fundamental application is to outsource the data to external cloud servers for scalable data storage. The outsourced data, need to be encrypted due to the privacy and confidentiality concerns of their owner. This results in the distinguished difficulties on the accurate search over the encrypted mobile cloud data. To tackle this issue, the searchable encryption for multi-keyword ranked search over the storage data. Specifically, by considering the large number of outsourced documents (data) in the cloud, utilize the relevance score and k-nearest neighbor techniques to develop an efficient multi-keyword search scheme that can return the ranked search results based on the accuracy.

Keywords- Cloud computing, multiple owners, multi-keyword search.

ARTICLE INFO

Article History

Received: 8th January 2017

Received in revised form :

8th January 2017

Accepted: 10th January 2017

Published online :

11th January 2017

I. INTRODUCTION

Cloud computing is gaining more popularity in the world. User can remotely store data on the server. With the advantage of flexibility and economic savings motivates both individuals and enterprises to outsource their local complex data management system into the cloud [2]. Cloud computing provides lots of benefits for user as well as enterprises to easy access, resource management, reduced cost etc. Despite of several benefits of cloud users are worried about the security for outsourced their data. These because once owner data is become outsourced, owner of the data completely lose control from the data. Virtualization and firewalls are security concerns supplied by the cloud service providers are not able to protect data privacy. In many researchers secure search over encrypted data has attracted the interest. This enables users for secure search without knowing the actual value of keywords and trapdoors. In this there are two protocols for different data owners use different keys to encrypt their files and keywords. Boolean keyword search scheme solves the problem of supporting efficient ranked keyword search. By doing this effective utilization of remotely stored encrypted data is achieved in Cloud Computing. It enhances system usability by returning the matching files [7]. This paper develop secure search protocol and proposed a novel Additive Order and Privacy Preserving Function family to protect the legal data from the attackers. Alexandra Boldyreva Nathan.

Chenette Adam O'Neill [18] addressed the problems of security of the ideal object" ROPF, for improving security of the any OPE (Ordr-Preserving Encryption) scheme. This system implements simple transformation that can work efficient to any OPE scheme. Efficiently Orderable Encryption (EOE), is proposed further for define general primitive of efficient OPE scheme. Dynamic secret key generation protocol and a new data user authentication protocol for preventing attackers from monitoring the secret keys and covering to be legal data [1].

Multi-Keyword Ranked Search:

To protect sensitive data from unwanted access i.e. provide data privacy. 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 for privacy and safe backup. On the other hand, the traditional 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.

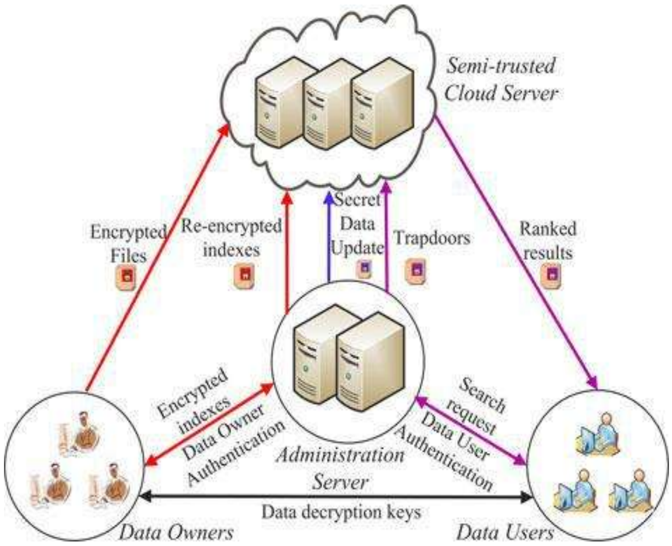

Fig. 1 System Architecture [1]

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 AND RELATED WORK

1. Secure and privacy preserving keyword search:

Qin Liu [3] proposed that the search provides keyword privacy, data privacy and semantic secure by public key encryption. CSP is involved in partial decipherment by reducing the communication and computational aerial in decryption process for end users. The user submits the keyword trapdoor encrypted by users‟ private key to CS (Cloud Server) securely and retrieve the encrypted documents.

Limitation: - The communication and computational cost for encryption and decryption is more.

2. Secure and Efficient Ranked Keyword Search:

Cong Wang [4] proposed search which solves processing overhead, data and keyword privacy, minimum communication and computation aerial. The data owner build index along with the keyword frequency based relevance scores for files. User request to cloud server by using the private key. The cloud server searches the index with scores and sends encrypted file based on ranked sequence.

Limitation: - It does not perform multiple keyword searches. Little overhead in index building.

3. Single Keyword Search Over Encrypted data on cloud:

Obtainable searchable encryption scheme consent to a user to firmly look for over encrypted data through keywords without first applying decryption on it, the proposed techniques support only conventional Boolean keyword search, without capturing any applicability of the files in the search result. When directly applied in large joint data outsourcing cloud environment, they go through next shortcoming.

Limitations:- Single-keyword search without ranking. Boolean-keyword search without ranking. Do not get relevant data.

4. Privacy-preserving Multi-keyword Text Search:

Wenhai Sun [6] proposed this search that provides similarity based search result ranking, keyword privacy, Index and Query confidentiality and Query Unlink ability. The encrypted file is built by vector space model supporting consolidated and distinctive file search. The searchable index is built using Multidimensional B tree. Owner creates encrypted query vector Q̄ for file keyword set. User gets the respective encrypted query vector of W from owner which is given to CS. Now CS searches index by Merkle–Damgård construction algorithm and compares cosine measure of file and query vector and returns top k encrypted files to user.

Limitation:- The similarity rank score of the document vector fully depends on the type of the document.

5. Secure Multi-keyword Top-k Retrieval Search:

Jiadi [7] proposed this search using Two round searchable encryption (TRSE). In 1st round, users submits multiple keyword 'REQ' ‘W’ as encrypted query for accomplishing data, keyword privacy and create trapdoor (REQ, PK) as Tw and sends to cloud server. Then cloud server calculate the score from encrypted index for files and returns the encrypted score result vector to user. In second round, user decrypt N with secret key and calculates the file ranking and then request files with Top k scores. The ranking of file is done on client side and scoring is done on server side.

Limitation: - The contraction and confining is used to reduce cipher text size, still the key size is too large. The communication aerial will be very high, if the encrypted trapdoor's size is too large. It does not make effective searchable index update.

6. Privacy Preserving Multi-Keyword Ranked Search (MRSE):

Ning [8] proposed this search for known cipher text model and background model over encrypted data providing low computation and communication overhead. The coordinate matching is chosen for multi-keyword search. They used inner product similarity to quantitatively evaluate similarity for ranking files. The drawback is that MRSE have small standard deviation σ which weakens keyword privacy.

Limitation: - Multi-keyword ranked search (MRSE) for known cipher text model may produce two different trapdoor which vague the privacy leakage problem of trapdoor unlink ability which may weaken the keyword privacy. MRSE has small standard deviation σ which in turn weakens the keyword privacy. The integrity of the rank order is not checked in MRSE.

7. Attribute-based Keyword Search:

Wenhai Sun [9] proposed Attribute-based Keyword Search that provides conjunctive keyword search; keyword semantic security and Trapdoor unlink ability. The owners creates index with all keywords and access list with policy attributes which specifies the users list authorized for searching. Now owners encrypt the document, index with access list using cipher text policy attribute based encryption technique. To have user membership management, they used proxy re-encryption and lazy re-encryption techniques to share the workload to CS. The user requests the Tw to CS using its private key. Now CS retrieves Tw and searches the encrypted indexes and return files only if the user’s attributes in Tw satisfies access policies in indexes which makes coarse-grained dataset search authorization.

Limitation: - Trapdoor generation will need more time with the increased number of attributes.

8. Efficient and Secure Multi-Keyword Search on Encrypted Cloud Data:

This proposed method has defined and solved the problem of effective but safe and sound rank keyword search over Encrypted cloud data [10]. Ranked search greatly enhances system usability by returning the matching files in a ranked order regarding to certain important criteria (e.g. keyword frequency) thus making one step closer towards sensible consumption of secure data hosting services in Cloud Computing. These papers has defined and solved the challenging problem of privacy preserving and efficient multi keyword ranked search over encrypted cloud data storage (MRSE), and establish a set of strict privacy requirements for such a protected cloud data utilization system to become a reality. The proposed ranking method proves to be efficient to go back extremely relevant documents corresponding to submitted search terms. The idea of proposed ranking method is used in our future system in order to enhance the security of information on Cloud Service Provider.

Limitation: - Dynamic updating and deletion of the document from the cloud is not possible.

9. A Secure and Dynamic Multi-keyword Ranked Search Scheme over Encrypted Cloud Data:

This proposed method [11] suggest a secure tree-based search scheme over the encrypted cloud storage, which supports multi keyword ranked search along 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 construction of index and generation of query to provide multi keyword ranked search output. To obtain 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 achieve 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 ensure accurate relevance score calculation between encrypted index and query vectors.

III. TECHNIQUES AND ALGORITHM

Some of the models, techniques and algorithms being used in the existing system are discussed and summarized as follows.

1. Vector Space Model

This model is used to represent the text by a vector of functions. The terms are the words and phrases. If words are considered as terms, every word becomes an independent dimension in a very high dimension vector space. If term represents a text, it gets a non- zero value in the text-vector along the dimension corresponding to the term. Text vectors are very space and no term is assigned a negative value.

2. Probabilistic Model

The principle of probabilistic model is that the documents in a collection should be ranked by decreased probability to query relevance. This principle is called as the probabilistic ranking principle. The ranking criterion is monotonic under log-odd transformations. Each probabilistic model that is proposed is based on a different probabilistic estimation technique.

3. Term Weighting

Term weighting is a technique that relies upon the better estimation of various probabilities. The main three factors play in term weight formulation is: Term Frequency - Words that repeat multiple times in a document.

Document Frequency - Words that appear in many documents are considered common. Document Length - When collection have documents of varying lengths, longer documents influence to score higher since they contain more words and more repetition.

4. Searchable Encryption Algorithm

An algorithm that consists of the polynomial time randomized algorithms. They are: KeyGen(s) - s is a security parameter taken and used to generate a key pair either public or private.

PEKS (Apub, w) - Apub is a public key and w is a word which are used to produce a searchable encryption.

Trapdoor (Apriv, w) - Apriv is a private key and w is a word which are used to produce a trapdoor Tw.

5. Cipher text Security

It is a technique that is used to provide security for the encrypted data. A cipher text attacker could easily break semantic security by reordering the keywords and submitting the resulting cipher text for decryption. A standard technique is used to break this and this technique is called the cipher text security.

6. Private Key Searchable Encryption

A model called private key searchable encryption is used to search on a private key encrypted data. The user himself encrypts data, so as to organize in an arbitrary way.

7. Public Key Searchable Encryption

Public key searchable encryption is a model that allows user to encrypt data and send it to the server. The owner provides decryption key may be different.

IV. CONCLUSION AND FUTURE WORK

In this survey paper, we have summarized different kind of searching techniques for the encrypted data over cloud. A systematic study on the privacy and data utilization issues is covered here for various searching techniques. Some of the important issues to be handled by the searching technique for providing the data utilization and security are keyword privacy, Data privacy, Fine-grained Search, Scalability, Efficiency, Index privacy, Query Privacy, Result ranking, Index confidentiality, Query confidentiality, Query Unlink ability, semantic security and Trapdoor Unlink ability. The limitations for all the searching techniques mentioned in this paper are discussed as well. From the above survey, we can say that security can be provided by the Public-Key Encryption and data security cam be provided by some different methods like fuzzy keyword search or can provide by binary balanced tree as an Index.

REFERENCES

[1] Wei Zhang, Student Member, Yaping Lin, Sheng Xiao, JieWu, Fellow, and 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] 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

[3] Privacy Preserving String Pattern Matching on Outsourced Data, Bargav Jayaraman

[4] Qin Liuy, Guojun Wangyz, and Jie Wuz, “Secure and privacy preserving keyword searching for cloud storage services”, ELSEVIER Journal of Network and computer Applications, March 2011

[5] Cong Wang et al., “Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data”, IEEE Transactions on parallel and distributed systems, vol. 23, no. 8, August 2012

[6] International Journal of Computer Applications (0975 – 8887) Volume 126 – No.14, September 2015

[7] Wenhai Sun et al., "Privacy-Preserving Multi-keyword Text Search in the Cloud Supporting Similarity-based Ranking", the 8th ACM Symposium on Information, Computer and Communications Security, Hangzhou, China, May 2013.

[8] Jiadi Yu, Peng Lu, Yanmin Zhu, Guangtao Xue, Member, IEEE Computer Society, and Minglu Li,” Toward Secure Multi-keyword Top k Retrieval over Encrypted Cloud Data”, IEEE Journal of Theoretical and Applied Information Technology 10th August 2014. Vol. 66 No.1 © 2005 - 2014 JATIT & LLS. All rights reserved. ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195 64 Transactions on dependable and secure computing, vol. 10, no. 4, July/August 2013

[9] Ning Cao et al., “Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data”, IEEE Transactions on parallel and distributed systems, vol. 25, no. 1, jan 2014

[10] Wenhai Sun et al., "Protecting Your Right: Attribute-based Keyword Search with Fine-grained Owner-

enforced Search Authorization in the Cloud", IEEE INFOCOM 2014, Toronto, Canada, April 27 - May 2, 2014

[11] Secure Ranked Keyword Search over Encrypted Cloud Data, IEEE PAPER, 2010.

[12] Sudha et al., “A Survey on Encrypted Data Retrieval in Cloud Computing”, International Journal of Advanced Research in Computer Science and Software Engineering 5(1), January - 2015, pp. 895-899

[13] Zhangjie Fu, Member, IEEE, Xingming Sun, Senior Member, IEEE, 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

[14] C. Wang, N. Cao, K. Ren, and W. Lou, Enabling Secure and Efficient Ranked Keyword Search Over Outsourced Cloud Data, IEEE Trans. Parallel Distrib. Syst., vol. 23, no. 8, pp. 1467-1479, Aug. 2012.

[15] H. S. Rhee, J. H. Park, W. Susilo, Trapdoor security in a searchable public-key encryption scheme with a designated tester," Journal of Systems and Software, vol. 83, no. 5, pp. 763771, 2010.