A Survey on Multi-Keyword Ranked Query Search over Encrypted Cloud Storage

Jaikishan Tindwani[1], **Aruna Gupta**[2]

[1]M.E. (Computer) Department of Computer Engineering, Jawantrao Sawant College of Engineering, Pune, India.
Savitribai Phule Pune University, Pune, Maharashtra, India-411007

[2] M.E. (Computer), Associate Professor , Department of Computer Engineering, Jawantrao Sawant College of Engineering, Pune, India.
Savitribai Phule Pune University, Pune, Maharashtra, India-411007

***Abstract:** Advancement in cloud computing have revamped the view of modern information technology which is motivating the data owners to outsource their data to the public cloud server like Amazon, Microsoft Azure, Google Drive, etc. With the help of data outsourcing, the organizations can provide reliable data services to their users without any concerns for the data management overhead. One more advantage of outsourcing the data over cloud as SAAS (Storage as a Service) is its cost-effectiveness, scalable and it can be accessed from anywhere and anytime. Normally, CSPs (Cloud Service Providers) take care of the data and its privacy, but there are some of the factors because of which the data privacy and user identity may be violated like an apostate employee, etc. Therefore, data owners should encrypt their respective sensitive data before outsourcing it to the public cloud server. Because the data is getting encrypted before outsourcing which may affect the performance of some important data accessing operations like searching of a document, etc. As we know CSPs plays a vital role for data privacy, but is it sufficient for the sensitive data like account figures, budgeting data, photos, health care files, etc? So, to answer the question, there are some methods/solutions offered to provide security and privacy to the data over cloud server. In the survey few of the searching techniques have been studied to find an effective method/solution for the retrieval of data/files over the encrypted cloud data.*

Keywords: Cloud Service Providers (CSPs), Cloud Storage, Ranked-Search, Encrypted-Data Search, secure cloud outsourcing **.**

1. Introduction

Currently we are in an information-explosion era where constantly purchasing new hardware, software and training IT professional is becoming a nightmare for almost every IT person. Coincidentally, we are witnessing an enterprise IT architecture which shifted to a centralized, more powerful computing paradigm known as Cloud Computing, in which enterprise‟s or personage‟s databases and applications are moved to the servers in the large data centers (i.e. the cloud) managed by the third-party cloud service providers (CSPs)a in the Internet. Cloud computing has been recognized as the most momentous turning point in the development of information technology during the past decade. People are attracted by the benefits it offers, such as personal and flexible access, on-demand computing resources configuration, considerable capital expenditure savings, etc. Therefore, many companies, organizations, and individual users have adopted the cloud platform to improve their business operations, research, or everyday needs.

With the remunerative option of pay-as-you-use, general and private data are outsourced by many individual users and organizations to third party CSPs. 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:

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

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.

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.

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 sensitive data might be like patient‟s medical records, financial data, etc. So, outsourcing plain data will raises some privacy issues. The data owner cannot afford to leak the private data to the CSPs or any unauthorized party. So, for such data owners, their data needs to be encrypted before outsourcing to the third party CSPs. Although 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 momentous for the search service to allow keyword query technique and provide result similarity ranking to meet the practical data retrieval as needed.

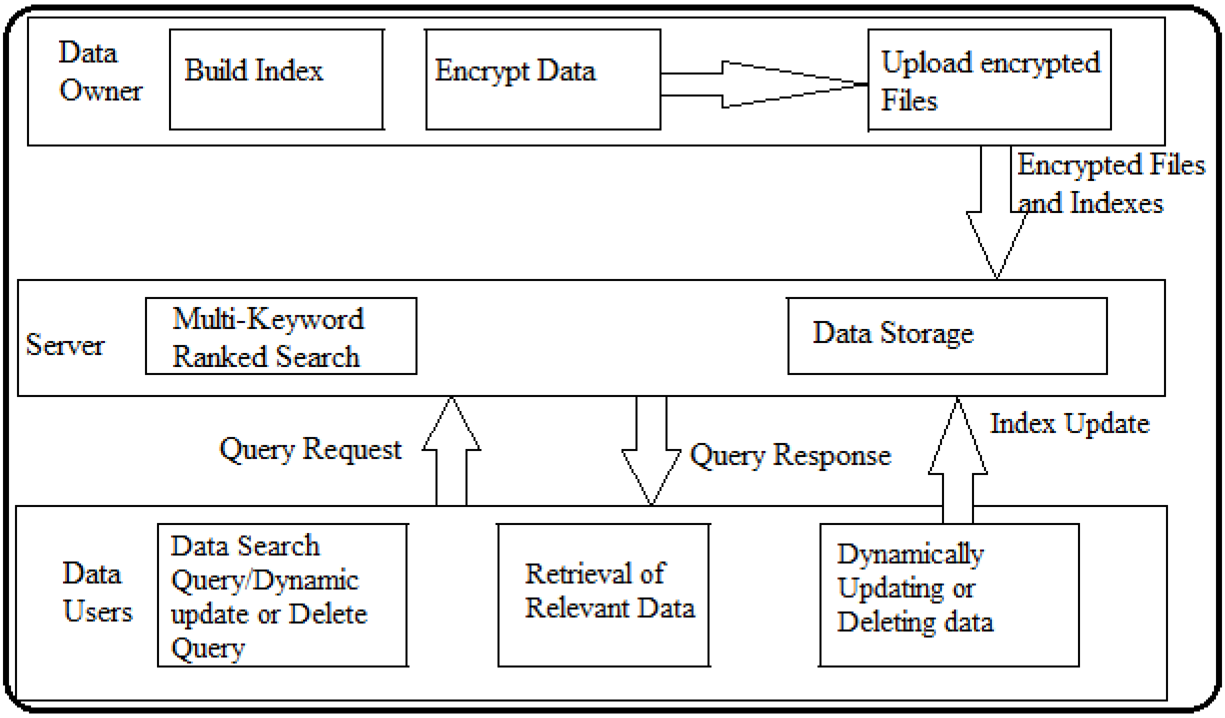
Data Owner
Build Index
Encrypt Data
Upload encrypted Files
Encrypted Files and Indexes
Server
Multi-Keyword Ranked Search
Data Storage
Query Request
Query Response
Index Update
Data Users
Data Search Query/Dynamic update or Delete Query
Retrieval of Relevant Data
Dynamically Updating or Deleting data

Figure 1: System Architecture

2. Literature Survey and Related Work

2.1 Secure and privacy preserving keyword search:

Qin Liu [3] proposed in this paper that the search that 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

2.2 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 „w" to cloud server with optional „k" as Tw 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

2.3 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..

2.4 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

2.5 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.

2.6 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.

2.7 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 ciphertext 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.

2.8 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.

2.9 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.

3. Techniques And Algorithm

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

3.1 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.

3.2 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.

3.3 Inference Network Model

A model that is used for a document to instantiate a term. The credit from multiple terms is accumulated given to compute the equivalent of a numeric score for data.

3.4 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.

3.5 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.

3.6 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.

3.7 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.

3.8 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.

4. 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 binary balanced tree as an Index

References

[1] Sheridan, J., Cooper, C.: Defending the cloud. http://www.reactionpenetrationtesting.co.uk/Defending%20the%20Cloud%20v1.0.pdf (2012)
[2] Privacy Preserving String Pattern Matching on Outsourced Data, Bargav Jayaraman
[3] 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
[4] 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
[5] International Journal of Computer Applications (0975 – 8887) Volume 126 – No.14, September 2015
[6] Wenhai Sun et al., "Privacy-Preserving Multikeyword Text Search in the Cloud Supporting Similarity-based Ranking", the 8th ACM Symposium on Information, Computer and Communications Security, Hangzhou, China, May 2013.
[7] Jiadi Yu, Peng Lu, Yanmin Zhu, Guangtao Xue, Member, IEEE Computer Society, and Minglu Li,”Toward Secure Multikeyword 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
[8] Ning Cao et al.,” Privacy-Preserving MultiKeyword Ranked Search over Encrypted Cloud Data”, IEEE Transactions on parallel and distributed systems, vol. 25, no. 1, jan 2014
[9] Wenhai Sun et al., "Protecting Your Right: Attribute-based Keyword Search with Finegrained Owner-enforced Search Authorization in the Cloud", IEEE INFOCOM 2014, Toronto, Canada, April 27 - May 2, 2014
[10] Secure Ranked Keyword Search over Encrypted Cloud Data, IEEE PAPER, 2010.
[11] Zhihua Xia, “A Secure and Dynamic Multi-keyword Ranked Search Scheme over Encrypted Cloud Data”, IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL: PP NO: 99 YEAR 2015
[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

Author Profile

Jaikishan Tindwani currently pursing M.E. (Computer Engineering) from Department of Computer Engineering, Jawantrao Sawant College of Engineering, Pune, India. Savitribai Phule Pune University, Pune, Maharashtra, Pune-411007. He received his B.E. (Information Technology) Degree from Prof. Ram Meghe Institute of Technology and Research, Badnera, Amravati, Maharashtra, India. Sant Gadge Baba Amravati University, Amravati-444607, Maharashtra, India

Aruna Gupta M.E. (Computer), Associate Professor , Department of Computer Engineering, Jawantrao Sawant College of Engineering, Pune, India. Savitribai Phule Pune University, Pune, Maharashtra, India-411007. She is awarded with the degree of B.E (Computer) and M. E (Computer).She has around 10 to 12 years of teaching and industrial Experience. She guided many students for the dissertation. She has published many national and international journals in this domain also. Her Research area includes Network Security and WSN.