International Journal of Management, IT & Engineering
Vol.15 Issue 9, September 2025
ISSN: 2249-0558 Impact Factor: 7.119
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gate as well as in Cabell's Directories of Publishing Opportunities, U.S.A

A Review Paper on Multi Keyword Search in Encrypted Cloud Data

Rajesh Dawar[1] , Manju Papreja[2*], Rashmi Chhabra[*]
[1]Govt College Barot,Gohana,Sonepat
[2,3]GVM Institute of Technology and Management, DCRUST, Sonipat, India
[1]gvmitrajesh@gmail.com
[2]manju.papreja@gmail.com
[3]rashmidahra@gmail.com
DOI: ijmra.ijmie.00930.33982

Abstract: Cloud Computing offers huge advantage to organizations and individuals to store their data onto the cloud and thus saves cost related to hardware and maintenance of data. As a result, users are prompted to put their data on to the cloud and retrieve it whenever needed from anywhere and using any device supporting internet. As data onto the cloud can be assessed by the cloud service provider, so data specially the sensitive data needs to be encrypted before putting the same onto the cloud. However, though the encrypting the data makes it safe but creates the problem of searching the files on the cloud containing specific keywords as server cannot make a plaintext search on encrypted data. But for efficient retrieval of data, we need keyword search functionality. To maintain user's data confidentiality, the keyword search functionality should not leak any information about the searched keyword or the retrieved document. This is known as privacy preserving keyword search. This paper put emphasises on design of privacy preserving keyword search over encrypted cloud data.

Keywords: Cloud Computing, Encryption, searchable encryption, data confidentiality.

INTRODUCTION

Cloud computing is a system where computer resources specially storage and computing power are made available to the user on demand and without direct management by the user. In essence, cloud computing is a construct that allows us to access applications that resides at a location other than the computer or other internet connected device. The beauty of cloud computing is that another company hosts your application and they handle the cost of servers, manage the software updates, and depending on your contract you pay for the service. The growth of high-speed networks, low-cost storage devices and low-cost computers has led to the widespread adoption of cloud. The great flexibility of cloud and cost savings resulting from it is prompting individuals and organisations to put their data onto the cloud. The sensitive data or the confidential data though for security reasons may

need to be encrypted before putting them onto the cloud. The same needs to be decrypted whenever the data on the cloud is retrieved by data user. However, if the data user needs to search something on the cloud it is quite infeasible to retrieve all the files, decrypt it and then finding out whether the file is the same the data user was looking for. In addition, the number of data users need not be limited to one, there can be n number of data users. Thus, storing data in encrypted form on the cloud serves limited purpose unless some facility is there to search for some data in the encrypted data files. The search criteria can be of multiple keywords; hence the search facility should provide for multi keyword search like the search facilities provided by modern search engines like Google. If the keywords which the data users are looking for appears in many files, some sort of ranking needs to be done and presented to the data user, so that he can retrieve the top k files needed and need not download all the files and thus waste network bandwidth. Some care also needs to be taken of misspelled words entered by user in search query, entering of matching word etc so that appropriate search can be done on the encrypted cloud data.

Literature Review

Ning Caoy. et. al. focussed on Boolean search [5]. This technique had two major drawbacks. In this user have to process each and every returned file to find the desired keyword. Second search process returns back all the files which are only dependent on presence or absence of query keywords. This increases network traffic and consumes bandwidth.

Li Wang. et. al. proposed wild card based fuzzy construction set (WFSC) [3] to enable fuzzy search over encrypted cloud data. The key concept was to maintain an index that covers all possible variations of keywords with in predefined edit distance. WFSC expands each extracted keyword into a set of modified keywords by inserting wild card character into the keyword. The drawback of this scheme is that with increase in edit distance the search quality is improved but there is a huge increase in modified keyword set.

Hui Cui et. al. proposed an expressive public key searchable encryption scheme [1] in the prime order groups, which allows keyword search policies i.e. predicates, access structures to be expressed in conjunctive, disjunctive or any monotonic Boolean formulas and achieves significant performance improvement over existing schemes.

Zhihua Xia Member IEEE et. al. presented a secure multi keyword ranked search scheme over encrypted cloud data [8] which simultaneously supports dynamic update operations like deletion and insertion of documents. Particularly, the vector space model and widely used TF – IDF model are combined in the index construction and query generation. Then a special tree-based index structure is constructed and a Greedy Depth first Search Algorithm is used

to provide efficient multi keyword ranked search. The secure kNN Algorithm is utilized to encrypt the index and query vectors, which ensures accurate relevance score calculation between encrypted index and query vectors. In order to resist statistical attacks, phantom terms are added to the index vector for blinding search results. Due to use of special tree-based index structure, the proposed scheme can achieve sub-linear search time and deal with the deletion and insertion of documents flexibly.

Priya S et al. showed a protected multi-watchword positioned look plan over scrambled cloud information [6], which all the while bolsters dynamic overhaul operations like cancellation and insertion of reports. In particular, the vector space model and the generally utilized TFIDF model are consolidated as a part of the record development and question era. They build an uncommon tree-based list structure and propose a "Ravenous Profundity first Seek" calculation to give productive multi-catchphrase positioned look. The protected kNN calculation is used to encode the record and question vectors, and in the interim guarantee exact importance score count between scrambled file and question vectors. Keeping in mind the end goal to stand up to measurable assaults, ghost terms are added to the record vector for blinding indexed lists. Because of the utilization of our uncommon tree-based list structure, the proposed plan can accomplish sub straight pursuit time and manage the erasure and insertion of reports adaptably.

K. Koussalya implemented watermarking approach to hide default pattern into image.[2] Water mark bits are embedded into image. So unauthorized users only get watermark data only. Based in inverse Discrete Wavelet Transform (DWT), we will get the seen water mark that can be restored into customary image. In the interface aspect, we will exchange the colour of textual content pixels into colour of photograph pixels. Person can set privacy settings to dam the pictures to down load by way of third parties. So unauthorized users most effective get watermark information handiest. Then utilizing disable options in mouse right click on and print reveal options. Snapshot privacy is maintained in social networks. Then using disable options in mouse right click and print screen options. Image privacy is maintained in social networks. Furthermore, the concept of blacklists and their administration are not believed by any of these access control models. The application of content-based filtering on messages posted on OSN user walls poses additional challenges given the short length of these messages other than the wide range of topics that can be discussed. Short text categorization has acknowledged up to now few attentions in the scientific community. This classifier will be used in hierarchical strategy. The first level task will be classified with positive and negative labels. The second level act as a negative, it will

develop gradual membership. This grade will be used as succeeding phases for filtering process. Short text classifier includes text representation, machine learning based classification.

Ning Cao, Member, IEEE, Cong Wang et al. addressed the problem of multi keywork ranked search over encrypted cloud data.[4] Among various multi-keyword semantics, he chose the efficient similarity measure of "coordinate matching," i.e., as many matches as possible, to effectively capture the relevance of outsourced documents to the query keywords, and use "inner product similarity" to quantitatively evaluate such similarity measure. For meeting the challenge of supporting multi-keyword semantic without privacy breaches, he proposed a basic idea of MRSE (Multi ranked search over encrypted Data) using secure inner product computation.

Xiaofeng Ding et al. focused on improving the efficiency and the security of multi-keyword top-k similarity search over encrypted data [7]. At first, we propose the random traversal algorithm which can achieve that for two identical queries with different keys, the cloud server traverses different paths on the index, and the data user receives different results but with the same high level of query accuracies in the meantime. Then, in order to improve the search efficiency, we design the group multi-keyword top-k search scheme, which divides the dictionary into multiple groups and only needs to store the top-ck documents of each word group when building index. Next, to protect the query unlink ability, we apply the random traversal algorithm to get the RGMTS, which can increase the difficulty of cloud servers to conduct linkage attacks on two identical `queries, and we can also tune the value of E to make the level of query unlikability flexible for data owners. Finally, the experimental results show that our methods are more efficient and more secure than the state-of-the-art methods.

Searchable Encryption

This section discusses about searchable encryption. We will discuss about the architecture, security requirements and approaches for design of searchable encryption schemes.

Architecture of Searchable encryption

The searchable encryption scheme uses the searched keyword to generate search token to enable the cloud server to retrieve the encrypted documents containing the searched keywords. The search token represents the encrypted query which can be generated by users with secret key. The architecture of searchable encryption comprises of four entities as given below:

- Data owner: The data owner is the individual or the organization which generates, encrypts the data, and puts it onto cloud server. The encryption is done using cryptographic algorithm which enables searching capability.

- Data user: The entity which sends encrypted query to CSP to search for documents containing specific keyword is termed as data user. Data user and data owner can be same entity or can be different entities.
- Cloud Service Provider (CSP): The entity which provides the facility for storage of documents and its retrieval to its users if termed as CSP.
- Key Generator: The trusted third-party entity which manages the generation and management of encryption and decryption keys.

Searchable encryption security requirements.

- The following requirements should be satisfied by a good searchable encryption algorithm:
- Retrieved data: Server should not be able to distinguish between documents and determine search contents
- Search query: Server should not learn anything about the keyword being searched for. Given a token, the server can retrieve nothing other than pointers to the encrypted content that contains the keyword.

• Query generation: Server should not be able to generate a coded query. The query can be generated by only those users with the relevant secret key.

• Search query outcome: Server should not learn anything about the contents of the search outcome.

• Access patterns: Server should not learn about the sequences and frequency of documents accessed by the user.

• Query patterns: Server should not learn whether two tokens were intended for the same query.

Design approaches for searchable encryption

Searchable encryption scheme can be implemented in two broad ways:

1) Non-keyword-based approach: In this entire document is scanned to find whether the keyword to be searched appears in the document or not.
2) Keyword-based approach: in this the index is built which contains each keyword of interest along with documents containing the specific keywords.

As is obvious that non-keyword-based approach takes a long search time for many documents. As against this keyword-based approach provides faster results for large set of documents. Hence our proposed technique will use keyword-based search scheme.

Proposed Solution

This paper aims to design a privacy preserving data retrieval and data storage system in cloud computing. This paper involves the use of searchable encryption algorithm to search for specific keyword in encrypted documents without requiring the user to download and decrypt the documents from the cloud server. The proposed technique will delegate the searching to CSP while maintaining privacy preserving approach i.e cloud server will not be able to infer any information about the keyword to be searched or any relation between the searched query and the documents being returned.

As Symmetric encryption algorithms takes less computational overhead as compared to asymmetric algorithms therefore, we will use the symmetric encryption algorithm for our purpose. Also, Keyword based approach will be used i.e. index will be built for the keywords to be searched and outsourced to CSP which will use this index to search for the keywords in a set of documents.

For our purpose user encrypts a set of Documents $D=\{D_1,D_2,\ldots\}$ and creates a encrypted index file I which contains a set of encrypted keyword which have been extracted from document set D. The user then uploads the document set D and index file I onto cloud server. Any user who wants to retrieve documents containing specific keywords creates an encrypted query and sends it to CSP. The cloud server uses the index to locate for the files containing specific keywords which are then returned back to the user.

The implementation of the said scheme can be broken into following modules:

1. Index Creation: Firstly, from the set of documents D, the keywords are extracted and an index is built. For our purpose each document is assigned a document id e.g. if we have 10 documents in our document set each can be assigned an id from 1 to 10. Thus, first document will have id as 1, second will have id as 2 and so on. Now the index will be built which lists the documents along with all the keywords appearing in the document. For instance, index will look like this:

INDEX

Document	ID	Keyword
D_1	1	w_1,w_3
D_2	2	w_1,w_2,w_9
D_3	3	w_3,w_9

Table -1

2. Creating inverted index: The next step will be to create inverted index from Table-1 which shows the document Ids for each keyword as shown in Table-2

Inverted Index

Keyword	ID
w_1	1,2
w_2	2
w_3	1,3
w_9	2,3

Table -2

3. Updating Inverted Index: To enable synonym based search i.e to allow for searching based on the synonym of the word entered by user e.g. institute word can be substituted for college and is the synonym of the same, so if user enters institute in the search query, then documents containing the keyword college should also be retrieved. To enable the same, the entry of the word should be copied in the inverted index table for synonym words. E.g. if word w_1 is college appearing in document 1 and 2, and w_7 is the word for institute so in inverted index w7 entry should be made to be appearing in document 1 and 2 as shown in table-3.

Updated Inverted Index

Keyword	ID
w_1	1,2
w_2	2
w_3	1,3
w_9	2,3
w_7	**1,2**

Table -3

4. Index Encryption: Now the index is encrypted using the key generated in step 1. The keyword encryption is computed as $ENC_{K1}(w_i||n_i)$ where ENC_{K1} represents encryption with key K_1, w_i is the keyword i and n_i is corresponding document id containing keyword w_i. For each of the encrypted keywords the encrypted index table lists the corresponding document Id.
5. Document Encryption: This encrypts each document with key k_2 and stores in database.
6. Search token generation: This step will be used by the user to generate a trapdoor/search token to be used at the server to perform a search over encrypted documents. For this user will input a search keyword and then compute the search

token. The search token for keyword w_z is computed as ENC_{K1} (w_z ||1), ENC_{K1} (w_z ||2),…. ENC_{K1} (w_z ||n) where n is the total number of documents. Once computed the same can be sent to the server to retrieve documents containing the search keyword.

7. Search: This module uses the search token and encrypted index as input and outputs the document list containing the specific keyword. For this it compares the search token with the encrypted index table to find the document ids containing the search keyword. It then sends the corresponding encrypted documents back to the client.
8. Decryption: the client decrypts the documents returned by the server using key K_2 to get the unencrypted documents containing the specified keyword.

Conclusion

As we entrust our data onto other servers, the need to maintain privacy of data increases. If the users are storing data on servers without any encryption, the information becomes public and hence becomes vulnerable to various kinds of security attacks. Users who want to keep their data safe from malicious attacks and wants to maintain the confidentiality and integrity of data encrypts their data before putting onto other servers. In this paper, a novel technique of searching keywords in encrypted files have been designed which also allows for synonym based search. The techniques designed can be used to search for the keywords in encrypted files without decrypting the files and enabling the cloud to search the keywords in encrypted files without even aware of the contents of the files and keywords and thus maintaining the privacy of data from various kinds of security attacks.

References

[1]. Hui Cui, Z. W. Efficient and Expressive Keyword Search Over Encrypted Data in Cloud. *Transcations on Dependable and Secure Computing 15,(3)*, 409-422 Research Collection School of Information System.

[2]. K.Kousalya, M. S. (2018). Image and Comment Privacy using Watermarking and Text Classification in OSN Framework. *International Journal of Engineering, Research & Technology (IJERT)*.

[3]. Li J.Wang, Q. C. (2010 proceedings IEEE). Fuzzy Keyword Search over Encrypted Cloud Data in Cloud Computing. *Info Comm* , 1-5.

[4]. N.Cao, C. M. (2014). Privacy Preserving Multi Keyword Ranked Search over encrypted Cloud Data. *IEEE Transcations on Parallel and Distributed Systems, Vol 25*.

[5]. N.Cao, C. M. (April 2011). Privacy Preserving Multi-Keyword ranked search over encrypted cloud data. *IEEE Infocomm*, 829-837.

[6]. Priya S., A. R. (Vol 5 Special Issue 10 , May 2016). A Multi-Keyword Ranked Search Scheme that is Dynamic and Secure over Cloud Data. *International Journal of Innovative Research in Science, Engineering and Technology*.

[7]. Xiaofeng Ding Member IEEE, P. L. (2016). Privacy Preserving Multi Keyword Top-k Similarity Search over encrypted Data. *IEEE Transcations on Dependable and Secure Computing*.

[8]. Zhihua Xia Member IEEE, X. W. (Vol 27 Feb 2016). A Secure and Dynamic Multi Keyword Ranked Search Scheme over Encrypted Cloud Data. *IEEE Transcation on Parallel and Distributed Systems*.

[9]. Md Iftekar Salam et al. (2015).Implementation of Searchable symmetric encryption for privacy preserving keyword search on cloud storage.