International Journal for Modern Trends in Science and Technology
Volume: 03, Issue No: 07, July 2017
ISSN: 2455-3778
http://www.ijmtst.com

An Efficient Multi-Keyword Search Scheme using Logical Operations in Cloud Environment

Karthic.D[1] | Kuppusamy.K[2]

[1]M.Phil., Scholar in Department of Computer Applications, Alagappa University, Karaikudi, Tamilnadu, India.
[2]Professor, Department of Computer Science, Alagappa University, Karaikudi, Tamilnadu, India.

To Cite this Article
Karthic. D and Kuppusamy.K, "An Efficient Multi-Keyword Search Scheme using Logical Operations in Cloud Environment", *International Journal for Modern Trends in Science and Technology*, Vol. 03, Issue 07, July 2017, pp. 77-80.

ABSTRACT

Cloud computing is growing incredibly, in recent era, in effective manner. Because of its huge storage capability, data owner can feed the information as an outsourcing and cloud server provides that data for the users on demand. While the public user is utilizing the data through the cloud server, there found lot of issues. The proposed method uses multi keyword search scheme to improve search scheme over the cloud. This proposed scheme has three major tasks, first is to generate the relevance score based on the term-frequency; second task isto accomplish the index depends on the correlated words from the content by using the classified sub-dictionaries technique. This sub-dictionaries technique is applied to attain higher potency and to enhance quick accessing of cloud data. Finally, the multi keyword search scheme consists of Boolean techniques such as AND, OR, NOT is applied. By using this technique the information can be retrieved from dataset in robust manner.

KEYWORDS: *Cloud Computing, Multi-keyword Search, Term Frequency, Index and Trapdoor Generation, Classified Sub-dictionary Technique.*

I. INTRODUCTION

In cloud computing, large amount of data can be stored and retrieved. The data owner can upload the data in cloud server. Data owner upload some data on the cloud. This data can be outsourced by various users and by accessing proper searching mechanism. Complexity of searching increased with increase in the amount of data stored. Keywords are used to search particular information needed. Hence, this work focused on multi-keyword search scheme for the data resided in cloud.

1.1 Multi keyword search scheme

Multi keyword search methods allow the user to retrieve files from cloud data. The cloud contains large amount of files and restricted user proficient to perform keyword search scheme. Text search scheme which considers the relevance scores of keywords based on vector space model[1],which is represented the total weight-age of term frequency occur in a files and display high performance factor terms as ranked[2].

A. Term Frequency and Inverse Document Frequency (TF & IDF)

Term frequency represents high score of main words in files. Inverse Document frequency to score the main of words in a document based on how frequently they appear across multiple documents.

B. Sub Dictionary Generation

Sub dictionary contains a collection of files that are indexed and represented as the IDF. In IDF, index buildings are generated as high weight-age terms are listed as one by one based on the relevance score and logical operation [4].

C. Trapdoor Generation

Trapdoor generation is provided by the cloud server for proper authentication purpose[3].The trapdoor key is randomly generated to provide the public user. By using the key,the public user can access the file from the cloud server.

II. LITERATURE REVIEW

The following literature review, consists of methods applied by the researchers worldwide and their recent developments in searching keyword

In [1] author **Zhihua Xia,** presented the secured multi-keyword ranked search scheme over the encoded cloud data, which is represented by the dynamic update operation such as insertion and deletion in the documents. The vector space model, Term Frequency ×Inverse Document Frequency techniques are joined to produce the index and query generation. Greedy depth first search algorithm is based on a special tree based index structure to present the well-organized multi-keyword ranked search. The safe kNN algorithm is used to encode the index and query vectors, while the extract accurate relevance score are estimated between the encoded index and query vectors. Advantages of this work are to use the special tree-based index structure.

In [2] authors **W. Sun,** et.al**,** presented theprivacy-preserving multi-keyword text search (MTS) scheme used similarity-based ranking over encoded data incloud. In multi-keyword search, search result ranking to construct the search index based onterm frequency and the vector space model with cosine similarity measure to suggest higher search result accuracy. The tree-based index structure and different adaptivemethods for multi-dimensional (MD) algorithm are developed for the purpose of, increase the search efficiency. They enhanced the search privacy scheme to construct the twosecure index schemes to meet the difficult privacy requirements under strong threat models is called cipher textand background model. Advantage of this scheme depends upon the index tree structure to enable authenticitycheck over the returned search results.

In [3] authors **H. Li, D. Liu, K. Jia, and X. Lin,** presented an authorized and ranked multi-keyword searchscheme (ARMS) over encrypted cloud data by included the cipher text policy attribute-based encryption(CPABE)and SSE techniques. Authors demonstrated on the searchable encryption technique that allows the search user to search over the encrypted data in cloud. The symmetric searchable encryption (SSE) technique is focused by this author. However, they do not conceive the search authorization problem that requires the cloud server to return the search results to authorized users.Security analysis demonstrates that the ARMS scheme can achieve confidentiality of documents collusion resistance and trapdoors unlink ability.

In [4] authors **D. X. Song, D. Wagner, and A. Perrig,** presented a well-organized privacy-preserving multi keywords search method over encrypted cloud data by used min hash functions. A multi-keyword search technique can be acombination of several keywords in a single query. By increasing the search constraints and also fetched the most relevant items returned to the search user. Since a multi-keyword search method that returns the matching encrypted data in a ranked ordered manner and it can hold three steps, first is to present a min hash based privacy-preserving multi keyword search method that provides high precision rates. Second step is to provides the security requirements and formally prove that it satisfies adaptive semantically security. Third stepis to use a ranking method depends on term frequencies and inverse document frequencies of keywords and demonstrate that it is efficient and effective by providing the implementation results.

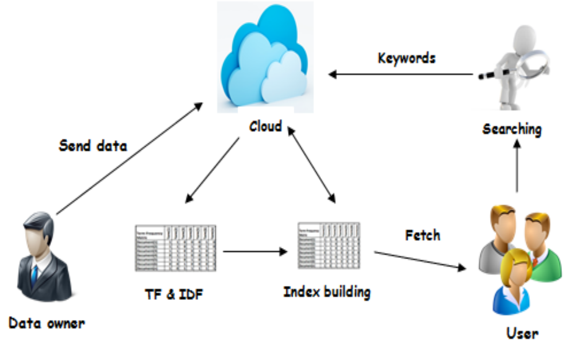


Figure 1 Architecture of the Efficient Multi-keyword Search scheme (EMS)

III. EFFICIENT MULTI-KEYWORD SEARCH SCHEME MODEL(EMS MODEL)

In this paper efficient multi keyword search method is proposed based on improved vector space model.

EMS, cloud server consists of huge amount of files both in encrypted and non-encrypted format. Searching over the cloud data remains as a major task. This research work focus on this, and provide a new efficient multi-keyword search model, for quick and ease of access of files in cloud. Our goal is to discover n word which occurred most in the file.Get all the word and store in the Hash-map as key and value would frequency of the word, if word is already present in the Hash-Map then increment the count (t). Once processing of the file is done then traverse through the hash map and return the k words with maximum counts.

Extracted the values and form aTF table. The values are applied to the TF(t)and IDF(t). The output of TF(t) and IDF(t) are multiplied and thenproduce a result are listed and sorted in largest manner. Highest values are placed on the sub dictionaryand filter by the required keyword files.

TF(t) = (Number of times words t appears in a document) / (Total number of words in the document)

IDF(t) = log_e (Total number of documents / Number of documents with words t in it).

TF-IDF = TF * IDF

Walk 1: let us consider, any type of textual data (*.txt, *.doc, *.java, *.pdf, etc.,) to be uploaded to the cloud (By using the AMAZON cloud) by the data owner. The owner shares the file to public user. The cloud contains huge amount of file.

Walk 2: Enter one keyword as the input of the searching textbox. Get the all type of textual files from the cloud and calculate the term frequency (Table-1). Extracted frequency values are put inTF(t) formula and form a tabular fashion(Table-2).

TABLE-1TF-Value

	java	Key	object
Doc1	50	60	20
Doc 2	70	0	19
Doc 3	0	0	0

Total No of words doc1 =1200
Total No of words doc2=1500

TF(1,1)=50/1200=0.041
TF(2,1)=70/1500=0.046

TABLE-2TF- Weight-age Value

	Java	Key	Object
Doc1	0.041	0.05	0.016
Doc 2	0.046	0	0.012
Doc 3	0	0	0

Walk 3: According to the result of TF (Table-1) which can applied in IDF(t). TF(t) and IDF(t) values are multiplied which is represented as TF-IDF(Table-3).

IDF(java)=log(3/2)=0.176IDF(key)=log(3/1)=0.477
IDF(object)=log(3/2)=0.176

TABLE-3TF-IDF Value

	Java	Key	object
Doc1	0.007216	0.02385	0.002816
Doc 2	0.008096	0	0.002112
Doc 3	0	0	0

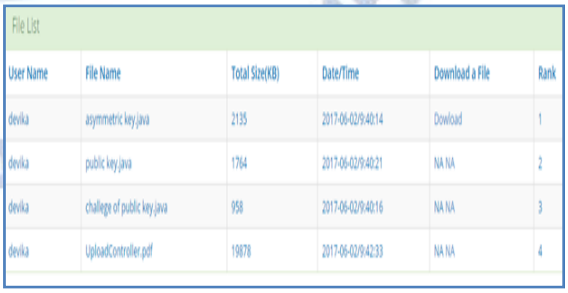
Walk 4: High score data can be retrieved and form a sub dictionary (Table-4).

TABLE-4High Score data are listed

1	Doc2	Java	1
2	Doc1	Java	2
3	Doc1	Key	1
4	Doc1	object	1
5	Doc2	object	2

Walk 5: Create an index block by using logical operation (Table-5).

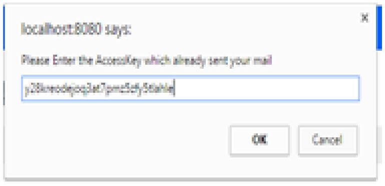
TABLE-5
Lists the file using the keyword as java, key, etc.



File List

User Name	File Name	Total Size(KB)	Date/Time	Download a File	Rank
devika	asymmetric key.java	2135	2017-06-02/9:40:14	Dowload	1
devika	public key.java	1764	2017-06-02/9:40:21	NA NA	2
devika	challege of public key.java	958	2017-06-02/9:40:16	NA NA	3
devika	UploadController.pdf	19878	2017-06-02/9:42:33	NA NA	4

Walk5: Download the file by using trapdoor key. Trapdoor function createsa random key for every time. It consists of both number and character sequence (figure 2).



localhost:8080 says:

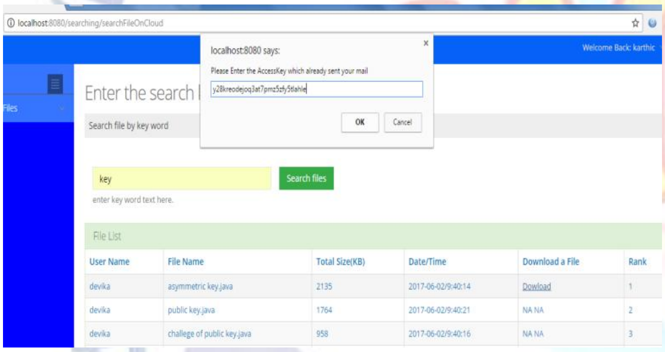
Please Enter the AccessKey which already sent your mail

y28kreodejoq3at7pmz5zfy5tlahle

Figure 2 –Trapdoor function generate key

IV. RESULT AND DISCUSSION

For multi keyword search method, the proposed method works efficiently. This method has content based searching keyword so it provides more accurate data. Trapdoor key is used to provide authentication to data. Multiple keywords are employed to search a single query depends it the TF and IDF. By using techniques of logical operation like AND, OR, NOT are reduced the Query complexity. Outcome of the result is more relevant. The experiment result takes less computation time for searching and listing files.



localhost:8080/searching/searchFileOnCloud

Welcome Back: karthic

localhost:8080 says:

Please Enter the AccessKey which already sent your mail

y28kreodejoq3at7pmz5zfy5tlahle

Enter the search

Search file by key word

key

enter key word text here.

File List

User Name	File Name	Total Size(KB)	Date/Time	Download a File	Rank
devika	asymmetric key.java	2135	2017-06-02/9:40:14	Dowload	1
devika	public key.java	1764	2017-06-02/9:40:21	NA NA	2
devika	challege of public key.java	958	2017-06-02/9:40:16	NA NA	3

Figure 3 - lists the files based on the relevance score

V. CONCLUSION

Cloud environment has a huge amount of data bundle. EMS scheme, which could help to find the files from the cloud server for extracted the keywords on files. Enhanced scheme classified sub-dictionaries helps to improve efficiency.The EMSrealizes secure and competent search with realistic functionality, like “AND”, “OR” and “NOT” operations.Analyze the protection of thatscheme in terms of confidentiality of documents and privacy protection of index and trapdoor generation.In that multi keyword search method has an advantage of data more accuracy and fast access.

References

[1] Zhihua Xia. [2016] “A Secure and Dynamic Multi keyword Ranked Search Scheme over Encrypted Cloud Data” IEEE transactions on parallel and distributed systems, Volume: 27, Issue: 2, Feb. 1 2016 **Page(s):** 340 – 352.

[2] W Sun, B Wang, N Cao, M Li, W Lou, YT Hou, H. Li. [2014],“Verifiable privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking”,*IEEE Trans. Parallel DistribSyst*, 25(11):3025–3035.

[3] H. Li, D. Liu, K. Jia, and X. Lin. [2015], “Achieving authorized and ranked multi-keyword search over encrypted cloud data”, in *Proc IEEE* **DOI:** 10.1109/ICC.2015.7249517.

[4] DX Song, D Wagner, A Perrig.[2000],“Practical techniques for searches on encrypted data”, in Proc. S&P, IEEE, pp. 44–55.