ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 2, Issue 3, March 2013

Improving the Efficiency of Ranked keyword Search over Cloud Data

Kiruthigapriya Sengoden, Swaraj Paul

***Abstract*— In this paper we define and solve the effective yet secure ranked keyword search over encrypted cloud data. We used order preserving symmetric encryption to protect the cloud data. Even though there are lots of searching techniques available, they are not giving efficient search results. For example the search results returned 40 records and in those 30 records are relevant and the remaining 10 records result contains irrelevant data. This paper mainly focuses on searching methods which will improve the efficiency of searching. We used both keyword search and concept based search methods in order to retrieve the relevance search criteria. This method will retrieve the documents based on broader conceptual entities, which will improve the efficiency of ranked keyword search.**

***Index Terms*— Concept Search, Confidential data, Cloud computing, Keyword Search, Order-preserving mapping, Ranked Search.**

I. INTRODUCTION

In Cloud Computing, data owners may share their outsourced data with a large number of users, who might want to only retrieve certain specific data files. One of the most popular ways to do so is through keyword-based search. Such keyword search technique allows users to selectively retrieve files of interest and has been widely applied in plaintext search scenarios. This existing searchable scheme will support only Boolean keyword search, which will combine words and phrases using the words AND, OR, NOT operators. This leads to following drawbacks,

1) Non relevant data search result
2) Large unnecessary network traffic, which is absolutely undesirable in today's pay-as-you-use cloud paradigm.
3) Decrease the efficiency and File retrieval accuracy.

So this kind of plaintext search method fails for cloud data. In order to improve the efficiency of ranked keyword search we used concept based searching techniques for file retrieval in which search words are conceptually related to the topic.

Manuscript received Feb, 2013.

***Kiruthigapriya Sengoden**, Computer Science, Vels University.,. Chennai, India,*
***Swaraj paul C,**Computer Science, Vels University., Chennai, India,*

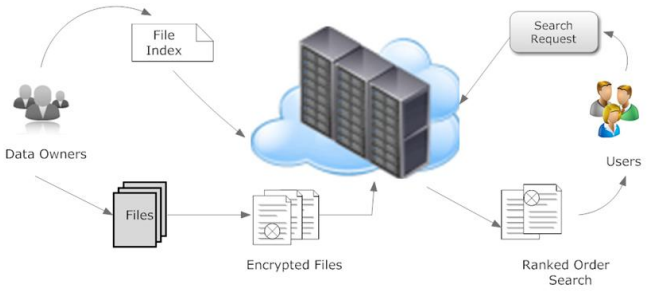

Fig1. Architecture Diagram

In above diagram data owner collect the files and will create index file for each of the file and then will encrypt the file using Order Preserving Symmetric Encryption algorithm and then store the data in cloud. In other end user will search for a data in cloud, the searched content will be encrypted format and we need to decrypt the data and then will display the search result in a ranked format. Ranked format of results are obtained using the score calculation of keywords, which is calculated based on document and term frequency algorithm and will be explained in detail in Section III. And also we used concept based search along with this, which greatly improves the efficiency of ranked keyword search. Finally the output by search result will contain relevant data as well as ranking of the word and frequency of the word will be displayed in a ranked format.

Then, we provide the framework of efficient ranked keyword and concept search system models Section 2, followed by Section 3, gives the data protection mechanism. And Section 4 on relevance score analysis evaluates the efficiency of search. Related work gives the different searching techniques comparative is discussed in Section 5. Finally, Section 6 gives the concluding remark of the whole paper.

II. FRAMEWORK OF EFFICIENT RANKED KEYWORD AND CONCEPT SEARCH

A) Setup

The data owner collect the data files and encrypt the files using OPS encryption and generate a secret key (refer Section 3 for more details). Then data owner generates the searchable index terms from the unique words which was extracted from file collection. The below table contains the sample words and index terms which was extracted from file collection.

Word	File Index
Soap	121
Software	132
Protocol	121
Dotnet	182

Table1. Index for words

Then the index terms are published on cloud server with encrypted file for the identification of files easily.

B) Score Calculation:

Once file indexing over, next we calculated the scoring. For calculating the score for each file, term frequency, document frequency, the length of files and the number of documents that the data owner has in his collection needs to be measured. The term frequency calculated based on how many times the keywords occurs in the same document , and for each file , and for each term this needs to be calculated. The document frequency calculated based on how many times a particular keyword exists in the different documents.

Score = (1/file length) * (1 + log (TermFrequency)) * (1+ log (No of Document / DocumentFrequency))

Given below table we obtained based on above score calculation formula.

Keywords	Term Frequency	Document Frequency	Score
Soap	3	0	0.00169
Software	6	1	0.00193
protocol	2	2	0.00169
Programming	3	1	0.00169
Dotnet	5	1	0.0026

Table2. Score calculation

The below diagram contains the list of keywords as x axis and term and document frequency in y axis .

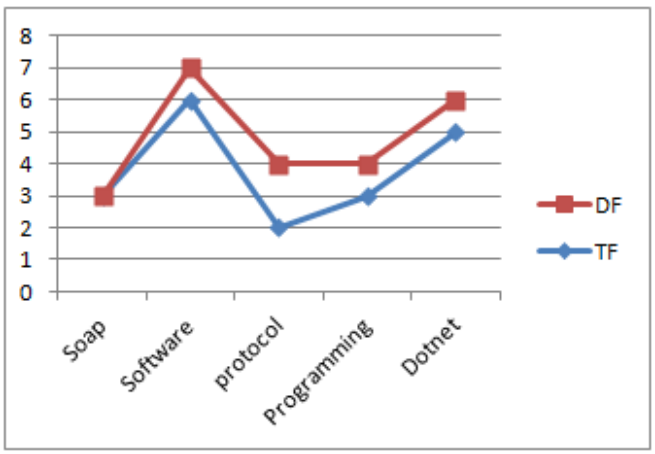

Fig2. Keywords Vs TF & DF

C) Ranked Keyword and Concept Search

In the file retrieval process user has not given the keyword directly as per his own suggestion, instead of that he sends the search request in the form of trapdoor generation. Before giving search request, user needs to aware about the index terms of file collections in the cloud server. Hence user requests the cloud server to view the index terms. The index terms are published for each and every file collection separately by the data owners during data outsourcing. Normally in cloud server, data user accesses the files after the authentication and authorization against the data owners and cloud server for data security. Once the user request the file to cloud server through a keyword, cloud server search the index terms related to the requested keyword and display the relevance results.

We initially used keyword search method to get the relevance results, which is effective when the exact search operations are conducted without any spelling errors. But if the proper keyword was not given by user means in that case the search engine fails to deliver satisfactory results. The Keyword search does not have the flexibility to conduct searches on a wide spectrum of terms. Then we added concept based searching techniques .This concept based search return a list of files that not only contain the exact search terms, but also search words are conceptually related to the topic, which provides a wider search scope capability.
So the combination of both keyword search along with concept search produce the relevant search result which greatly improve the efficiency of search.

III. DATA PROTECTION

In order to protect the sensitive cloud data in this concept based ranked keyword search we used order preserving symmetric algorithm to encrypt/decrypt the documents. The given below diagram explains the concept of OPS,

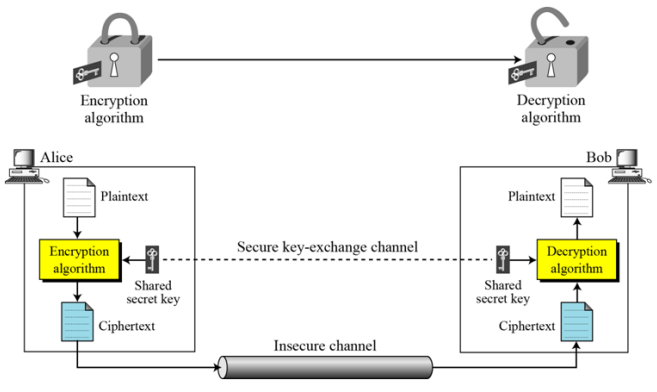

Fig3. Order preserving symmetric Encryption

For example, considered Alice and Bob needs to share the information , in that case if they going to share the plain text format then easily anyone can access or able to view the files.

In order to protect the cloud data we used OPSE which is secure channel of exchanging the data. The above diagram

explains the security mechanism that we used, first Alice contains the plain text which will be converted into encryption format and then shared secure key will be generated by OPSE and based on that key Bob can authenticate into application and will decrypt the plaintext which will be in a human readable format.

IV. Relevance Score Calculation

We conducted a thorough experimental evaluation of the proposed techniques on real data set. The performance of our scheme is evaluated based on relevance score distribution using concept based searching methods.

The efficiency of our proposed one to- many order-preserving mapping is determined by both the size of score domain M and the range R. The below diagram explains the efficiency measurement of our proposed scheme. The two figures show the value distribution after one-to-many mapping with as input the same relevance score set of keyword "network," but encrypted with two different random keys. The result represents the mean of 100 trials.

A) Efficiency of Search System

The efficiency of search system calculated based on relevance scoring. We done this analysis for set of keywords and calculated score based on their relevance and provided ranked search results.

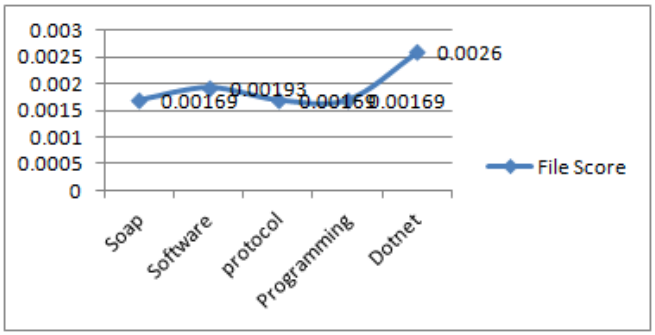

Fig4. Keywords Vs Relevance scoring

The above diagram contains list of keywords as x axis and file score in y axis. We achieved this using concept based searching methods which greatly increase the search results efficiency.

V. Related Work

We have done a survey on existing searching methods (PTSED - Practical Techniques for Searches on encrypted data ,APKS - Authorized Private Keyword Search over encrypted data, SI – Secure Index and PKE – Public key encryption) and summarized with following Characteristic: 1) Sequential approach: This method will find a particular keyword in a document, which will check for every one of its elements, and will display the search result one at a time and in linear order and this will decrease the performance i.e for example Searching "a[a-z]b" ,needs 26 queries. 2) Document Index: Storing a secure keyword index in cloud. This kind of index will allow a query to check if the documents contain a keyword and will retrieve the files. It will not search for the entire document based on index keyword will retrieve the documents that are especially useful for large documents and large document. 3) Perform keywords filter: Indexing of keyword contains unique keywords, it will not contain the duplicate keywords in index files.4) Public Key authentication: This kind of encryption will allow anyone to access the data in cloud, which is not efficient one.

Characteristic	PTSED	APKS	SI	PKE
Sequential Approach	Yes	No	No	Yes
Document Index	No	Yes	Yes	No
Keyword Filter	No	Yes	Yes	No
Public Key	Yes	No	No	Yes

Table3.Comparative Study on existing search methods

We have done a comparative study on existing searching methods with above factors and found a efficient way to improve the ranked keyword search. We used searching method which contains ,document indexing , performing keyword filter using concept based searching and order preserving symmetric encryption schema to protect the sensitive cloud data which will use private key encryption techniques will greatly improve the efficiency of ranked keyword search.

VI. Conclusion

In this paper, we proposed a searching method to improve the efficiency of ranked keyword search. We gave introduction about the existing searchable encryption framework, it is very inefficient to achieve efficient ranked search. We proposed a efficient one-to-many order preserving mapping function, which allows the effective RSSE to be designed. In additional to that we proposed combination of concept based and keyword based searching techniques .This kind of techniques has the ability to categorize, and search large collections of unstructured text on a conceptual basis. This kind of searching technique is more reliable and efficient search method that is more likely to produce relevant results than traditional searches. Our experimental relevance score analysis results show that the proposed search methods greatly improve the efficiency of ranked keyword search.

REFERENCES

[1] C. Wang, N. Cao, K. Ren, and W. Lou, "Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data" , Proc. IEEE , Parallel and Distributed Systems, Aug. 2012.
[2] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure Ranked Keyword Search over Encrypted Cloud Data," Proc. IEEE 30th Int'l Conf. Distributed Computing Systems (ICDCS '10), 2010.
[3]] E.-J. Goh, "Secure Indexes," Technical Report 2003/216, Cryptology ePrint Archive, http://eprint.iacr.org/, 2003..
[4] Dan Boneh,Giovanni Di Crescenzo,Rafail Ostrovsky,Giuseppe Persiano, "Public Key Encryption with Keyword Search" , 2007
[5] Xirong Li , "Harvesting Social Images for Bi-Concept Search" , Proc. IEEE , Multimedia, Aug. 2012
[6] D. Song, D. Wagner, and A. Perrig, "Practical Techniques for Searches on Encrypted Data," Proc. IEEE Symp. Security and Privacy, 2000.
[7] Ming Li, Shucheng Yu†, Ning Cao, Wenjing Lou, "Authorized private keyword searches over encrypted data"