ISSN 2395-1621

Multi-keyword Ranked Search Over Encrypted Cloud Data

[#1]Ingule Devaki Vijay, [#2]Jadhav Aditya Anand, [#3]Shende Swapnil Narayan, [#4]Pawar Shailesh Sharad, [#5]Bhagat Sangram Veersen

[1]devaki.ingule@gmail.com,
[2]adityajadhav575@gmail.com,
[3]swapshende007@gmail.com,
[4]pawar.shailesh93@gmail.com,
[5]bhagatsangram007@gmail.com

[#1]Assistant Professor, Department of Computer Science & Engineering, COE Phaltan, India.

[#2345]Students, Department of Computer Science & Engineering, COE Phaltan, India.

ABSTRACT

Now a days cloud computing has become more popular, so more information possessors are actuated to their information to cloud servers for great convenience and less monetary value in data management. However, sensible information should be encrypted before outsourcing for public. In this paper the problem of a secure multi-keyword search on cloud is solved by using encryption of data before it actually used. Which are continuously supports dynamic modify operation like insertion and deletion of the documents.

Keywords: Cloud Computing, Ranked based search, Download frequency, Multikeyword search, Encrypted cloud data, Synonym query.

ARTICLE INFO

Article History

Received: 15th March 2017

Received in revised form :

15th March 2017

Accepted: 17th March 2017

Published online :

22nd March 2017

I. INTRODUCTION

Cloud computing has become new model which handles large resources of computing. Services provided by the cloud computing is storage and on demand services, both the individuals and organizations are motivated to the cloud. Instead of purchasing software and hardware devices.

Cloud provides secure online storage and there is no loss of data, the data is available at anytime and anywhere. Paper shows the general approach for data protection is to encrypt the data by using AES algorithm.

The simple method for downloading data is decrypts it locally, because consumers want to search needed data rather than all. In this way it is essential to investigate a productive and successful search benefit over encrypted outsourced information.

The current search approaches like ranked search, multi-keyword search that empowers the cloud clients to locate the most pertinent information rapidly. It likewise decreases the system activity by sending the most important information to client asks. However, in genuine search situation it may be conceivable that client searches with the equivalent words of the predefined keywords not the correct keywords, because of absence of the client's correct information about the information.

II. LITERATURE SURVEY

Zhangjie Fu, Xingming Sun, Nigel Linge and Lu Zhou [2] proposed a successful way to deal with take care of the issue of multi-keyword ranked search over encrypted cloud information supporting synonym queries. To address multi-keyword search and result positioning, Vector Space Model (VSM) is utilized to build document index that is to state, each document is communicated as a vector where each dimension value is the Term Frequency (TF) weight of its comparing keyword. Another vector is additionally produced in the question stage. The vector has a similar dimension with document index and its each dimension value is the Inverse Document Frequency (IDF) weight. At that point cosine measure can be utilized to register comparability of one document to the search inquiry. To enhance search proficiency, a tree-based index structure which is an adjust paired tree is utilized.

C. Wang, N. Cao, J. Li, K. Ren, and W. Lou [3] developed the Ranked search that increases system usability by returning the relevant files in a ranked order.(e.g., keyword frequency). In this state-of-the-art searchable symmetric encryption (SSE) security definition used for increasing its

efficiency. They have also proposed the existing cryptographic primitive, order preserving Symmetric encryption (OPSE) for searching matching files.

W. Sun, B. Wang, N. Cao, M. Li, W. Lou, and Y. T. Hou [5] proposed a method to address the problem of similarity-based ranking is privacy-preserving multi-keyword text search (MTS) scheme. They also presented the search index based on the vector space model, i.e., cosine measure, and TF IDF weight to achieve high level of search accuracy and to support a multi-keyword queries with search ranking functionalities.

III.PROBLEM STATEMENT

Many associations and organizations store their more significant data in cloud to protect their information from infection and hacking. The advantage of new computing is it looks deeply for cloud clients. Rank search enhances framework ease of use by ordinary coordinating records in a ranked arrange with respect to certain importance criteria (e.g. - Keyword and download frequency). As straight forwardly outsourcing significance scores will trickles a great deal of delicate data against the keyword security, to solve this problem we proposed asymmetric encryption with ranking consequence of query information which will give just expected information.

IV. PROPOSED SYSTEM

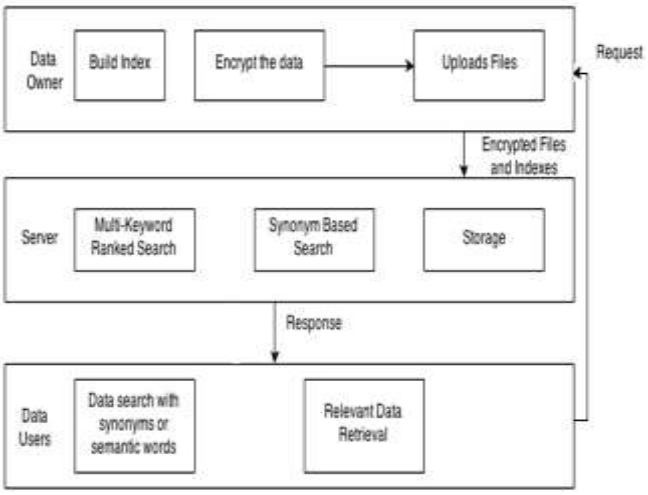

Fig. 1. System Architecture of Multi-Keyword Ranked Search Over Encrypted Cloud Data.

1. Keyword Expansion

To enhance the accuracy of search results, the keywords are removed from outsourced content documents required to be stretched out by regular synonyms or comparable words, as cloud customers, searching information may be the synonyms of the predefined keywords.

2. Upload Encrypted Data

After expansion of keywords the data owner assist data with encrypting the document utilizing AES Algorithm and after that upload the encrypted document to the cloud for storage reason. This permits data owner to store their secret key in extremely secure way without presenting it to the clients of framework. For this, secret key is put away again in encrypted frame.

3. Search Module

This module helps clients to enter their query keyword to get the most important documents from set of uploaded documents. This module recovers the documents from cloud which coordinates the query keyword.

4. Rank Generation

In data recovery, a positioning capacity is as a rule used to assess relevant scores of coordinating documents to a demand. The rank capacity in view of the term recurrence (TF) and converse document recurrence (IDF) is utilized as a part of expanded organize i.e. TF-IDF. Additionally this framework gives client most mainstream documents for their keywords by examining history of most downloaded documents for specific inquiry keywords.

5. Download Ranked Results

Clients can download the resultant arrangement of documents just if he/she is approved client who has allowed consent from data owner to download specific document. Owner will send encrypted secret key and session key to client to decrypt the document.

V. METHODOLOGIES

a. AES algorithm

AES is an iterative instead of Feistel cipher. It depends on "substitution–permutation network". It contains an arrangement of linked operations, some of which include supplanting inputs by particular yields (substitutions) and others include rearranging bits around (permutations). Strangely, AES plays out every one of its calculations on bytes instead of bits.

Steps for AES algorithm:

1. Create a random key for symmetric encryption of user facts.
2. Encrypt the records the use of this random key.
3. Encrypt the random key the use of asymmetric encryption.
4. Send the encrypted message and the encrypted key to the receiver of searched results.

Henceforth, AES treats the 128 bits of a plaintext obstruct as 16 bytes. These 16 bytes are organized in four columns and four rows for preparing as a matrix.

b. TF-IDF

TF:

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

IDF:

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

- $f_{d,j}$, the TF of keyword w_j within the document d;
- f_j, the number of documents containing the keyword w_j;
- M, the total number of documents in the document collection;
- N, the total number of keywords in the keyword dictionary;
- $w_{d,j}$, the TF weight computed from $f_{d,j}$;
- $w_{q,j}$, the IDF weight computed from N and f_j;

The definition of the similarity function is as follows:

$$SC(Q, D_d) = \frac{\sum_{j=1}^{N} w_{q,j} \cdot w_{d,j}}{\sqrt{\sum_{j=1}^{N} (w_{q,j})^2} \cdot \sqrt{\sum_{j=1}^{N} (w_{d,j})^2}} \quad (1)$$

Where $w_{q,j} = 1 + \ln f_{d,j}$, $w_{q,j} = \ln(1 + \frac{N}{fj})$ .The normalized *TF* and *IDF* weight are $\frac{w_{d,j}}{\sqrt{\sum_{j=1}^{N} (w_{d,j})^2}}$ and $\frac{w_{q,j}}{\sqrt{\sum_{j=1}^{N} (w_{q,j})^2}}$ respectively, and hence, the vector Q and D_d are both unit

VI. CONCLUSION

The multi-keyword ranked methodology results in the more effective search handle which diminishes the network traffic and download bandwidth. It gives back the precisely matched documents, and also the records which incorporate the terms semantically significant to the question keyword. It offers fitting semantic separation between terms to achieve the question keyword expansion. The encryption has been executed to ensure the security also, efficiency of information, before it is outsourced to cloud, and gives protection to datasets, indexes and keywords.

REFERENCES

[1] Zhihua Xia, Xinhui Wang, Xingming Sun, Qian Wang,"A Secure and Dynamic Multi-keyword Ranked Search Schema over Encrpted Cloud Data," IEEE Transactions On Parallel And Distributed Systems,Vol:PP No:99 ,Year 2015

[2] Zhangie Fu, Xingming Sun, 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

[3] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure ranked keyword search over encrypted cloud data," Proceedings of IEEE 30th International Conference on Distributed Computing Systems (ICDCS), pp. 253-262,2010.

[4] Q. Chai, and G. Gong, "Verifiable symmetric searchable encryption for semi-honest-but-curious cloud servers," Proceedings of IEEE International Conference on Communications (ICC"12), pp. 917-922, 2012.

[5] W. Sun, B. Wang, N. Cao, M. Li, W. Lou, and Y. T. Hou, "Privacy preserving multi-keyword text search in the cloud supporting similarity based ranking," ASIA CCS 2013, Hangzhou, China, May 2013, ACM pp. 71-82, 2013.

