Efficient Keyword Search Techniques over Cloud Data in Cloud Computing

Ms. Gayathri M R1
M.Tech Student
Dept. of ISE
Acharya Institute of Technology
Bengaluru, India

Prof K. Srinivas2
Assistant Professor
Dept. of ISE
Acharya Institute of Technology
Bengaluru, India

Abstract- **Cloud computing is becoming an essential part of IT industry. As it becomes dominant, more amounts of sensitive data are being outsourced into the cloud. Users stores their files in different Cloud models (Public, Community and Hybrid), and retrieves on need basis [1]. On the basis of user's requirement, accessing the data from cloud servers is one of the major challenging problems. Processing the query and retrieving the data over cloud server is one of the difficult tasks. Encryption of data is also an important concept that needs to be dealt in-detail from both Data Owner & Data User perspective. Many keyword based searching techniques are used for the purpose of retrieving the data from cloud. In this paper various data search techniques like Fuzzy keyword search, Single keyword search, and Ranked keyword search are concentrated to focus on accessing encrypted Cloud data.**

Keywords- Encryption, Keyword, Searching, Fuzzy, Wildcard, Gram

I. INTRODUCTION

Cloud computing is the method of retrieving the services with no knowledge of the exact physical location of the data. Cloud computing is mentioned as a Utility computing as it uses pay for usage model. Users have to pay only for their usages / cloud access period. Using the technology of cloud computing, users can access a various resources like stored programs, data storage and application development platforms. Cloud may be considered as extension of object oriented programming as it uses the concept of data abstraction. Now cloud services are highly scalable, which permit customers to pay only for the power and storage they have used. Moving the data from a highly secured data centre to Internet based cloud model will need more effect on security and privacy. Implications are needed for every type of services and applications which is hosted on cloud.

Data will be stored in Cloud spanning across different data centres across geographic locations. Data access / retrieval involve extracting only the required data from a cloud database. The retrieved data will be stored as a file, viewed or printed for additional actions. The ability to query and retrieve the data based on some user defined principles is an essential feature of the data storage and retrieval system.

Data will be stored as both public as well as private in the cloud. Different searching strategies are available for both the types of data. The personal data will be stored in the cloud using encryption technique. Only authorized users, with required privileges and knowledge of the access keys can retrieve the data. Accessing the data from encrypted storage is one of the major challenges that user's face.

Different types of searching techniques are used to search for encrypted data over cloud [2] [3].

In section II, a comparative study of different search techniques are provided. In section III, Multi-keyword search, Ranked Search and Semantic search techniques relevance to the current cloud search is elaborated.

II. COMPARATIVE STUDY OF DATA SEARCH TECHNIQUES

A. Fuzzy Keyword Search

The fuzzy keyword search method gives the search results based on the following rules:

- If the user's searching input accurately matches the pre-defined set of keywords, the server will gives the files having the keyword.
- If there are some errors in spelling or some format variations in the searching input, the server will gives the nearby possible results based on pre-specified similarity semantics.

Technique for Constructing Fuzzy Keyword Sets

- Wildcard Based Technique

Wildcard based technique allows users to search for all files that contain similar qualities. The technique can be useful if a user cannot remember a specific file name or would like to see an entire grouping of files that are created to share some part of their name.

For example, keyword CASTLE with pre-set edit distance 1, its wildcard based fuzzy keyword set can be constructed as,

CASTLE, 1= {CASTLE, ASTLE, C*ASTLE, C*STLE,,*CASTLE,*CASTL*E,CASTL……}

Edit Distance method can be performed in 3 ways:

1. Substitution: Changing one character to another in a word

2. Deletion: Deleting one character from a word

3. Insertion: Inserting a single character into a word

- Gram Based Technique

Another effective technique for creating fuzzy set is based on Grams. The gram of a string is a substring that can be used as a signature for efficient estimated search. While gram has been commonly used for creating inverted list for estimated string search, we use gram for the purpose of matching. Any simple edit operation will disturb one exact character of the keyword, leaving all the remaining characters untouched.

For example, for the keyword CASTLE, fuzzy keywords can be constructed using Gram Technique as:

{CASTLE, CSTLE, CATLE, ASTLE ,CASLE, CASTE, CASTL, …}

The obtained fuzzy keywords are nothing but the possible substrings of the given keyword CASTLE

Both wildcard based technique and gram based technique's goal is to construct fuzzy keyword sets. Wild card technique uses edit distance concept, whereas gram based techniques uses the substrings of a given string to obtain fuzzy keyword sets.

Advantages:

1. Maintains keyword privacy
2. Active consumption of remotely stored encrypted data

Disadvantages:

1. Needs large storage capacity as it generates so many possible fuzzy keywords from a given keyword
2. Cannot accommodate ranked search, multi-keywords search, and semantics-based search

B. *Single Keyword Searching*

In Single keyword search, as the name suggests, searching is carried out for matching the documents that contain one or more words specified by the user. This paper focuses on presenting a traditional single keyword searchable encryption schemes, usually built on Encrypted Searchable Index. Its contents are hidden to the server unless it is given appropriate trapdoors generated via secret key(s). But in this system anyone with public key can write to the data stored on server but only authorized users with private key can do searching.

For example, a file CASTLE.txt is present in cloud. If we want to access that file using the technique of Single keyword search, we have only one keyword. Consider ASTL is the specified single keyword, and when we do searching we will get more number of data files corresponding to the keyword ASTL. User may get BASTL.txt, DASTL.txt; ASTL1.txt etc. in their search result as all these file names contain ASTL keyword. Selecting CASTLE.txt among large data set will be difficult for users and also it consumes more time. Hence, single keyword searching is not efficient to do searching over large data set.

Advantages:

1. Easiest method to implement for data search

Disadvantages:

1. Support only exact keyword searching
2. Cannot be used to search over large data set

C. *Index Management Searching*

This method is a searchable re-encryption scheme. With this technique, user can share the data with others safely by creating searchable encryption index and re-encrypting it [5].

The security needs set up uses two techniques: Proxy re-encryption method and searchable encryption method. These methods provide efficiency. The search technique uses more than one keywords and thus the flexibility and efficiency is provided. But it is not very effective when compared to other search techniques like Multi-keyword, Ranked and Synonym methods.

Advantages:

1. Efficiency is calculated in terms of calculation volume
2. Traffic efficiency is provided by using one round of communication process for keyword search
3. This method affords quick search

Disadvantages:

1. Index collected from multiple keywords with variable length is not possible

III MULTI-KEYWORD SEARCH, RANKED SEARCH AND SEMANTIC SEARCH TECHNIQUE

A. *Multi-Keyword Search*

It is a type of searching technique that looks for the matching data files present in cloud that contain multiple words specified by the user. Because of multiple keywords search, user obtains the most suitable matches. The search result appears in the list, it allows the users to find the most appropriate data quickly [6].

For example, a file CASTLE.txt is present in cloud. If we want to access that file using the technique of Multi-keyword search, we have multiple keywords generated by this multi-keyword searching technique. Consider ASTL, STLE, CALE, ALES are some of the generated multiple keywords. We can do searching using any of these keywords, we will get more accurate data files in search result. Hence selecting CASTLE.txt among obtained search result set will be easy for users. Hence, Multi-keyword search technique is more efficient and effective when compared to other searching techniques.

Advantages:

1. Improves the search results accuracy
2. Reduces the network traffic

Disadvantages:

1. Does not support syntactic and semantic search

B. Ranked Keyword Search

This scheme sorts out the matches by relevance. The ranking appears in the search results list, it allows the users in finding the best related data quickly rather than arranging the every match in the content collection [4].

Ranked search greatly increases system usage by giving matching files in a ranked order regarding to certain relevance criteria, thus creating one step closer towards deployment of privacy-preserving data hosting services in Cloud Computing.

For example, a file CASTLE.txt is present in cloud. If we want to access that file using the technique of ranked search, we can use any of the specified keywords present in the filename. Consider ASTL is the keyword, and when we do searching we will get more number of data files corresponding to the keyword ASTL. Search result may contain more data files which contain the keyword ASTL. But search result will be in ranked order, i.e. data files with only ASTL keyword will be in the top list and files with similar filename will come in the next search results. Ranked search also generates more number of data files. But selecting CASTLE.txt among large data set will be not much difficult for users, as the search result comes in order. Hence, ranked keyword searching will be efficient, when it is used along with the multi-keyword method.

Advantages:

1. Displays the most relevant results, among less data files
2. The order of results retrieved is maintained
3. Reduces the network traffic

Disadvantages:

1. Efficient when it is used with multi-keyword technique

C. Semantic Search

Semantic search technique’s purpose is not only to find specified keywords, but it will also determine the similar meaning of the word which is used to search. Whenever the user inputs the synonyms of the predefined keywords, cloud server perform searching and gives data files which contain specific keywords [7].

For example, a file CASTLE.txt is present in cloud. If we want to access that file using the technique of semantic search, we have synonyms of the multiple keywords generated by the technique. Consider ASTL, STLE, CALE, ALES are some of the generated multiple keywords, semantic search will generate random synonyms to all the keywords. When we want to do searching, we can use both multiple keywords and the corresponding synonym keywords for best result. Using this method, we will get more accurate data files in search result. Hence selecting CASTLE.txt among obtained data set will be very easy for users, as the search results are less. Hence, Multi-keyword search with synonym technique is more efficient and effective when compared to other searching techniques.

Advantages:

1. Gives back more accurate search result
2. Requires less time to do searching when compared with other searching techniques

Conclusion

This paper focused on different data search techniques across cloud servers. In the beginning, need for cloud data storage and its importance are discussed. Later paper went through cloud data searching importance. After that we discussed about different searching techniques in the cloud data. Each technique has their advantages and disadvantages. In order to overcome the disadvantages, Multi-keyword searching scheme with synonym query is used.

References

[1] Deepa P L, S Vinoth Kumar, Dr S Karthik, "Searching Techniques in Encrypted Cloud Data", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 1, Issue 8, October 2012

[2] S. Balasubramaniam, Dr. V. Kavitha, "A Survey on Data Retrieval Techniques in Cloud Computing", Journal of Convergence Information Technology (JCIT), Volume8, Number16, November 2013.

[3] Chengyu Hu, Pengtao Liu," Public key encryption with ranked multi-keyword search*",* 5th IEEE International Conference on Intelligent Networking and Collaborative Systems, 2013

[4] Vimmi Makkar, Sandeep Dalal "Techniques of keyword search over cloud data A Review", International Journal of Computer Applications & Information Technology Vol. 3, Issue I June-July 2013 (ISSN: 2278-7720).

[5] S. Ramamoorthy, R. Saravanan,
" Sharing Secure Data in the Cloud for the Multiuser Group", International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 3, Issue 1, January – February 2014.

[6] Zhangjie 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.

[7] Prof. C. R. Barde, Pooja Katkade, Deepali Shewale and Rohit Khatale, "*Secured Multiple-keyword Search over Encrypted Cloud Data* ", International Journal of Emerging Technology and Advanced Engineering , Volume 4, Issue 2, February 2014

[8] Ms. Archana D. Narudkar, Mrs. Aparna A. Junnarkar,"A Survey on Searching Techniques over Encrypted Data*",* (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (2) , 1007-1010, 2015