Implementation of Ranked Search over Encrypted Cloud Data with Supporting Synonyms

Santhosh H¹  Ms. V.K. Annapurna²
¹²Department of Information technology
¹²National Institute of Engineering, Mysore

Abstract—“Implementation of ranked search over encrypted cloud data with supporting synonyms” has been developed to provide a stable and efficient search services in the cloud data storage. The cloud service provider will charge the cost for the services they provide to the user so consumers want to find the most relevant product, information or data, which is highly desirable in the “pay-as-you use” cloud computing paradigm. The searchable schemes over encryption data support only exact or fuzzy keyword search. That is, there is no tolerance of synonym substitution and/or syntactic variation. And the searchable schemes support only single keyword search. To meet the challenge of effective search system over secured data, the new search schema is described which is efficient and flexible searchable scheme that supports both multi-keyword ranked search and synonym based search i.e Semantics-based multi-keyword ranked search technology over encrypted cloud data which supports synonym queries is proposed. The search results can be achieved when authorized cloud customers input the synonyms of the predefined keywords, not the exact or fuzzy matching keywords. Possible synonym substitution and/or her lack of exact knowledge about the data to the data user. Ranked search over encrypted cloud data will protect the data in the cloud and retrieve the required document or page that contain search key efficiently and quickly which is very important in the pay-per-use type of cloud model, so it saves time and money. If the user doesn’t have knowledge of data in particular, he can also obtain the search result because search results can be obtained for synonym of input keywords.

Keywords: Fuzzy, CSP

I. INTRODUCTION

Consumers want to find the most relevant product, information or data, which is highly desirable in the “pay-as-you use” cloud computing paradigm. Consumer-centric cloud computing is a new model of enterprise-level IT infrastructure which provides on demand, high quality applications and services from a shared pool of resources for consumers. On the other hand, some of the problem can be caused by Cloud Service Provider (CSP) because they have full control of the outsourced data. There may exist unauthorized operation on the outsourced data on due to curiosity or for the profit.

And in the existing search approaches cannot accommodate requirements like ranked search, multi-keywords search, semantics-based search etc. The ranked search enables cloud customers to find the most relevant information quickly. Ranked search can also reduce network traffic as the cloud server sends back only the most relevant data. Multi-keyword search is also very important to improve search result accuracy as single keyword search often return coarse search results. In the real search scenario, it is quite common that cloud customer’s searching input might be the synonyms of the predefined possible synonym substitution such as commodity and goods, and/or her/his lack of exact knowledge about the data. The existing searchable encryption schemes support only exact or fuzzy keyword search.

To meet the challenge of effective search system over secured data ranked search over encrypted cloud data with supporting synonym is proposed. Search schema is described which is efficient and flexible searchable scheme that supports both multi-keyword ranked search and synonym based search.

To address multi-keyword search and result ranking the Vector Space Model (VSM) is used to build document index. Each document is expressed as a vector where each dimension value is the Term Frequency (TF) weight of its corresponding keyword. A new vector is also generated in the query phase. The vector has the same dimension with document index and its each dimension value is the Inverse Document Frequency (IDF) weight. Then cosine measure can be used to compute similarity of one document to the search query. To improve search efficiency, a tree-based index structure which is a balance binary tree is used. The searchable index tree is with the document index vectors. So the related documents can be found by traversing the tree.

A. Objective:
The main concern of this paper is to protect the data in the cloud and retrieve the required document or page that contain search key, efficiently and quickly. Which is very important in the pay-per-use type of cloud model, hence it saves time and money.

If the user doesn’t have exact idea of data that is stored in the cloud, they can also obtain the search result because search results can be obtained for synonym of input keywords.

B. Problem Statement:
To search the data over encrypted cloud efficiently, proposed a fuzzy keyword search scheme over encrypted cloud data, it address only the problems of minor typos and format inconsistence.

Second proposed a ranked search scheme, the cloud server can rank relevant data files with no knowledge of a specific keyword weight. But this scheme supports only single keyword search.

Third they proposed a ranked scheme supporting multi-keyword. It supports only accurate ranked search results.

This paper will add the synonym support for search key to achieve the multi-keyword ranked search in the encrypted cloud.
C. Existing System:
Existing search methods over the encrypted cloud are:
1) Fuzzy keyword search over encrypted data in cloud computing technique supports keyword with exact match.
2) Boolean search is a searchable encryption schemes that allow users to securely search over encrypted data through keywords, without capturing any relevance of data files.
3) Multi-keyword ranked search over encrypted cloud data allows multiple keywords in the search request and return documents in the order of their relevance to these keywords.

D. Drawbacks of Existing Systems:
1) The above Fuzzy keyword has no tolerance of minor typos and format inconsistencies.
2) Boolean search has to retrieve all files containing the queried keyword so it cause unnecessary network traffic, which is absolutely undesirable in today’s pay-as-you-use cloud paradigm.
3) Multi-keyword ranked search over encrypted cloud data doesn’t retrieve the document or file with synonyms of the entered key words (i.e. related meaning of the key word that is given for search key).

E. Proposed System:
Semantics-based multi-keyword ranked search technology over encrypted cloud data which supports synonym queries is proposed. Search results will be achieved for synonym of input keywords, due to the possible synonym substitution and/lack of exact knowledge about the data.
The proposed system involves mainly three key steps.
1) Synonym expansion, in order to improve the accuracy of search results, the keywords extracted from outsourced text documents need to be extended by common synonyms.
2) Rank function, in information retrieval, a ranking function is usually used to evaluate relevant scores of matching files to a request. It is done by identifying the term frequency and inverse document frequency.
3) Searchable Index Tree is a searchable index is a balance binary tree. Searching the document or the content for the entered key by the user is done by post order traversal of the index tree.

F. Methodology:
The development of the proposed system took many stages. In the first stage, the problem of existing system was studied. Then the analysis the project requirements were carried out. Further the system was developed stage by stage from requirement analysis to the implementation.

II. SYSTEM REQUIREMENT AND SPECIFICATIONS
A. Requirement Specification:
This section shows the functional requirements that are to be satisfied by the system. The entire requirement exposed here are essential, that is, a system would not be acceptable that does not satisfy some of the requirement presented here.

1) Software Specification:
   1) Operating system : windows XP/7
   2) Coding Language : ASP.net, C#.net
   3) Tool : Visual Studio 2010
   4) Database : SQL SERVER 2008

2) Hardware Specification:
   1) Processor : Pentium 4 and above.
   2) Ram : 512Mb and above.
   3) Hard Disk : 40 GB.
   4) Input device : Standard Keyboard and Mouse.
   5) Output device : VGA and High Resolution Monitor.

III. SYSTEM DESIGN PHASE
Design is the process of converting a user oriented description of the data into a computer based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system.

A. System Architecture/Model:

Fig. 3.1: System Architecture

System Architecture provides a high level overview of how the functionality and responsibilities of the system were partitioned and then assigned to subsystems or components. And it is described in the Figure 3.1. This architecture consists cloud server, data owner and data user.

In our system architecture as shown in fig 3.1

B. Data Owner:
1) He or she encrypts the document and stores the encrypted document in cloud server.
2) Select the distinct key words from the document before encrypting document. These key words are selected based on the term frequency (TF) and inverse document frequency (ITF).
3) Creates the tree based index for the key words and store it in the cloud server.

C. Data User:
1) He / she will make the search request by entering key words to the cloud server.
2) Cloud server sends result in the form of encrypted document to the data user.
3) Data user will ask the key to data owner to decrypt the document.
4) Data owner will send the key to data user to their respective email id’s.
5) Data user will decrypt the document.
D. Cloud Server:
1) Receives the search request from the data user in the form of keywords.
2) Cloud server will search these keywords that are entered by the data user in the tree based index that was created and stored by the data owner.
3) Cloud server will extract the synonym of the search key words.
4) Cloud server retrieve the all the document that contains synonym of the search key word through the tree index.
5) Send the encrypted document to the user.

IV. ADVANTAGES, DISADVANTAGES AND APPLICATIONS
A. Advantages:
1) This search approach reduces search time and avoids the irrelevant search results.
2) Saves the cost for the data user because the search result contains only the relevant document.
3) User who doesn’t have any idea about the data stored in the cloud can also obtain the search result because the search approach supports the synonym query.
4) Provides the security to the data because the document is encrypted before it is out source to the cloud.

B. Applications:
1) This type of search approach is applicable for the cloud application.
2) Can be deployed in daily applications such as medical field, business, education field etc. Because the important documents will be stored in the cloud server.

V. CONCLUSION
This paper, proposes an effective approach to solve the problem of synonym-based keywords search over encrypted cloud document. The main contributions are summarized in two aspects: synonym-based search and similarity ranked search. The search results can be achieved when authorized cloud customers input the synonyms of the predefined keywords, not the exact or fuzzy matching keywords, due to the possible synonym substitution and/or her lack of exact knowledge about the data.

The aim is that cloud consumers can search the most relevant products or data by using the search system.

REFERENCES
Websites referred:
1) 1.www.isi.edu
2) 2. www.ece.ubc.ca
3) 3.www.nsnam.org
4) 4. scholar.google.co.in