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. Traditional searchable encryption schemes allow a user to securely search over encrypted data through keywords without first decrypting it, these techniques support only conventional Boolean keyword search, without capturing any relevance of the files in the search result. When directly applied in large collaborative data outsourcing cloud environment, they may suffer from the following two main drawbacks. On the one hand, for each search request, users without pre-knowledge of the encrypted cloud data have to go through every retrieved file in order to find ones most matching their interest, which demands possibly large amount of post processing overhead On the other hand, invariably sending back all files solely based on presence/absence of the keyword further incurs large unnecessary network traffic, which is absolutely undesirable in today’s pay-as-you-use cloud paradigm.

Keywords :- RSSE , OPSE ,IAAS,PAAS,SAAS etc

1. Introduction

Today, the latest paradigm to emerge is that of Cloud computing which promises reliable services delivered through next-generation data centers that are built on virtualized compute and storage technologies. Consumers will be able to access applications and data from a “Cloud” anywhere in the world on demand. The consumers are assured that the Cloud infrastructure is very robust and will always be available at any time. Computing services need to be highly reliable, scalable, and autonomic to support ubiquitous access, dynamic discovery. In particular, consumers indicate the required service level through Quality of Service parameters, which are noted in SLAs established with providers. Of all these paradigms, the recently emerged Cloud computing paradigm appears to be the most promising one to leverage and build on the developments from other paradigms.

Cloud Layers:-
1. Infrastructure as a Service (IAAS)
2. Platform as a Service (PAAS)
3. Software as a Service (SAAS)

Infrastructure-as-a-Service (IAAS):
Infrastructure-as-a-Service(IAAS) like Amazon Web Services provides virtual servers with unique IP addresses and blocks of storage on demand. Customers benefit from an API from which they can control their servers.

Platform-as-a-Service (PAAS):
Platform-as-a-Service(PAAS) is a set of software and development tools hosted on the provider’s servers. Developers can create applications using the provider’s APIs. Google Apps is one of the most famous Platform-as-a-Service providers.

Software-as-a-Service (SAAS):
Software-as-a-Service (SAAS) is the broadest market. In this case the provider allows the customer only to use its applications. The software interacts with the user through a user interface. These applications can be anything from web based email, to applications like Twitter.

2. Types of Cloud
propose to bring together the advance of both crypto and IR community to design the ranked searchable symmetric encryption (RSSE) scheme, in the spirit of “as-strong-as-possible” security guarantee. specifically, we explore the statistical measure approach from IR and text mining to embed weight information (i.e., relevance score) of each file during the establishment of searchable index before outsourcing the encrypted file collection. As directly outsourcing relevance scores will leak lots of sensitive frequency information against the keyword privacy, we then integrate a recent crypto primitive order-preserving symmetric encryption (OPSE) and properly modify it to develop a one-to-many order-preserving mapping technique for our purpose to protect those sensitive weight information, while providing efficient ranked search functionalities.

3. Proposed System

Secure Ranked Keyword Search over Cloud Data

Development of a private cloud is very expensive. Storage of sensitive data in public cloud is very risky. To make it possible, unauthorized access is avoided by storing the data in encrypted format. This paper tackles the problems of enabling searchable encryption system with support of secure ranked search in order to implement the top k retrieval. In this paper, statistical measure approach from IR and text mining to embed weight information of each file during establishment of searchable index before outsourcing the encrypted file collection is explored.

Team frequency: Number of times a particular keyword appears within the file. Inverse document frequency (IDF): It is calculated as the total number of files by the number of files in particular keyword. Ranking function: It is calculated by using TF*IDF rule.

Algorithm for index table generation:
Step 1. Read the document F
Step 2. Segment the document term wise and encrypt with key
Step 3. Calculate term frequency (TF) and inverse document frequency(IDF) and publishing time(PT)
Step 4. Generate index table(Itable) and files upload to server.

Algorithm 1 Ranked Search

For all documents Ri do
Compare(level1 index of Ri , query index)
  j = 1
  while match do
    increment j
    Compare (levelj indices of Ri, query index)
    end while
  rank of Ri = highest level that match with query index
end

4. Related work:
Now-a-days cloud servers get to store large amount of files. Here select and processing the files is the main problem. Whenever large numbers of files are available in cloud server under encryption some problems are generated. Totally all files are not encrypted. Although traditional searchable encryption schemes allow a user to securely search over encrypted data through keywords without first decrypting it, these techniques support only conventional Boolean keyword search, without capturing any relevance of the files in the search result. That’s here there is no sufficient privacy and security in outsourcing.

Fig 2 Architecture of cloud Server

1. It can retrieve the results with less communication overhead.
2. It can provide the results with effective retrieval accuracy.
3. It can provide effective privacy and security application.

In the architecture we have three entities

Data owner: Data owner having collection of data files that he wants to outsource into the cloud server in encrypted format, this will increase effective data utilization.

Data user: When the data user wants to search the required files he enters a keyword in a secret form.

Cloud server: It is the place where a pool of data files and different applications can store. Previously user can select the files in the form of a plain text files. This is ailing under access the files. There is no perfect decryption technique to access the files of representation process. Here we introduce encryption based secure keyword searching mechanism. It can provide efficient solution for accessing the data. It is a good usability to display the effective matching details files. These matching files are extracted with relevance score. This kind of matching files are retrieved with efficient mechanism. It can provide the results with guaranteed mechanism. All the files are collected with encryption format. All encrypted files are given weight in implementation process. These kinds of approaches show the better result in implementation.

Design Goal
To make possible ranked searchable symmetric encryption for successful employment of outsourced cloud data under the mentioned model. Our system design should achieve the following security and performance assurance. Specifically we have the to reduce the size of index. A list of standard IR techniques can be adopted, including case folding, stemming, and stop words etc. We omit this process of keyword extraction and refinement and refer readers to for more details Ranked search. In order to rank the documents, a ranking function is required, which assigns relevancy scores to each document matching to a given search query. One of the most widely used metrics in information retrieval is the term frequency. Term frequency is denied as the number of times a keyword appears in a document. Instead of using term frequency itself, we assign relevancy levels based on the term frequencies of keywords.

To enable ranked keyword search for effective utilization of outsourced cloud data under the model, our system design should achieve the following security and performance guarantee.

a) Ranked keyword search: For efficient searching process the process use the mechanism of Topic detection and tracking. The search time includes fetching the posting list in the index, decrypting, and rank ordering each entry.

b) Security guarantee: For providing the security in the cloud server, this process uses the privilege method.

5. RESULTS AND DISCUSSION

Execution process:

a. Owner uploads the file in to the cloud server, and set the privilege to the particular user for easily access data.
b. And give the particular permission like write, read or both for providing the security. Here the user’s are separated by authorized user and unauthorized user. Authorized user is the owner permitted person and unauthorized user is unpermitted person.
c. So authorized user easily access the data from the cloud server by using the ranked efficient keyword search by the mechanism of TDT4 mechanism. Unauthorized user asks the permission to access the data.
d. After the data owner permission, then only the

<table>
<thead>
<tr>
<th>Number of files</th>
<th>Per keyword list size (KB)</th>
<th>Per list build time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>6.272</td>
<td>2.80s</td>
</tr>
</tbody>
</table>

Efficiency of the ranked keyword search:
This type of ranked keyword search enhances the efficient usage of outsourced files by providing Inter cloud communication constantly between data owners and users. So that the cloud server learn nothing from the data uploaded by data owners. The search time is not affected while fetching the posting list in the index, decrypting, and rank ordering each entry

BENEFITS
Ranked keyword search: to explore different mechanisms for designing effective ranked search.
a. Provide more security to the data owner, by means of Inter-cloud communication through e-mail.
b. Authentication of both search results and the outsource process enables search engine to be more robust than before in cloud environment
c. Privilege method is used for the security. So process has the more security compared to the existing system.

**Fuzzy Keyword Search**

This keyword search deeply enhances system usability by returning the matching files when users searching inputs accurately match the predefined keywords or the closest possible matching files based on keyword similarity semantics, when exact match fails. Usage of edit distance to quantify keywords similarity and development of a technique for construction of fuzzy keyword sets. Fuzzy eliminates the need for enumerating all the fuzzy keywords and the resulted size of fuzzy keywords sets is extensively concentrated.

**Goals of introducing fuzzy keyword search are:**

I) To discover new mechanism for constructing storage efficient fuzzy keyword sets

II) To design well-organized and effective fuzzy search scheme based on the constructed fuzzy keyword sets.

III) To validate the security of the planned approach.

![Fig 3 Data Retrieval in from the Cloud Server](image)

When a user wants to access the data by keyword search mechanism, first he wants to get the authorization. That is the authorized person can only retrieve the data. The authorization is provided by a key which is randomly generated. That key is unique for each user. The user should remember this key throughout the searching process. The user can enter the key words which is the conjunction of single keywords. That is AND, OR, BOTH and he get a search result which is in a ranked order.

In the existing system we are giving a conjunctive keyword for search and retrieve the data. Here an advanced tire-tree is used for storing this conjunction of keywords and searching each separately. The AND, OR, BOTH are also defined. We are using ‘gram based’ method and ‘wild card’ method for ‘fuzzy keyword construction’.

In both these methods the conjunction of keyword is implemented, which will produce a highly efficient ranked result.

**Main modules in Fuzzy keyword search are:**

a) **Wildcard-based technique:** To edit the operations at the same position a wild card based technique is used. We can calculate the edit distance by using substitution, deletion and insertion.

b) **Gram-based technique:** Here the fuzzy set is constructed based on grams. The gram of a string is a substring and it can be used for efficient estimated search. The order of the characters after the primitive operation is always kept the same before the operations.

c) **Symbol-based tire-traversed scheme:** In this technique, for storing the fuzzy keyword set over a finite symbol set, a multi way tree is constructed. Here we consider a cloud data system consisting of data owner, data user and cloud server. Given her a collection of n Encrypted data files C = (F1, F2, FN) stored in the cloud server, a predefined set of distinct keywords W = (w1, w2, we), the cloud server provides the search service for the authorized users over the encrypted data C. We assume that the authorization between the data owner and users is appropriately done. An authorized user types in a request to selectively retrieve data files of user’s interest. The cloud server is responsible for mapping the searching request to a set of data files, where each file is indexed by a file ID and linked to a set of keywords. The fuzzy keyword search scheme returns the search results according to the following rules: If the user’s searching input exactly matches the pre- set keyword, the server is estimated to return the files containing the keyword; if there are typos and/or format inconsistencies in the searching input, the server will return the closest possible results based on pre-specified similarity semantics.

6. **Conclusion:**

In this paper, we proposed a searching method to improve the efficiency of ranked keyword search Algorithms. 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**


Ranked Keyword Search over Outsourced Cloud Data”, IEEE transactions on parallel and distributed systems, vol. 23, no. 8.