ISSN:0975-9646

Techniques for Efficient Keyword Search in Cloud Computing

P.Niranjan Reddy,Y.Swetha

Department of CSE , Kakatiya Institute Of Technology & Science,
Warangal Dist-506002,India.

Abstract— As cloud computing becomes most general, the important information is centralized into the cloud server. To protect the data stored in the cloud, the data must be encrypted. Although traditional encryption techniques allows the user to securely search through the keyword and return retrieved files, these techniques are useful only for exact keyword search. In this paper, we solve the problem of exact keyword match by providing searching with fuzzy keyword. We also propose two more techniques called gram based technique which is useful for reducing the time, providing fast searching and increase the performance by considering substring from the given string. And Symbol-based tree traverse search scheme where a multi way tree structure is built by using symbols, which works for more than one keywords entered by the user. By providing security, we show that the proposed solution is secure and privacy-preserving.

Keywords- fuzzy keyword, gram, traversal search

I. INTRODUCTION

Cloud computing is the technology used to access remotely stored data through the internet. It protects the data from the disasters like earthquakes, tsunami, cyclones, fire etc. Cloud computing protects the data by using emails, personal records, documents, etc. By storing the useful data into the cloud, the owners are free from the burden of maintenance. In this, owners can share their data with the large number of users when the users request for the data. The benefits of cloud computing are burden-free for data owner, on demand service, independent of location and etc. The users might wants to retrieve only specific data files in which they are interested. To provide this facility, one of the best way is to selectively retrieve the files through keyword based search instead of providing all the information irresptive of users interest. Such keyword based search techniques provides only the data which the users wants such as google, Wikipedia etc. The traditional encryption schemes builds an index for each keyword and this index is associated with the files. This technique can be called as secure trapdoors. But it only works for ***exact*** keyword match which restricts the users to perform keyword search which is not suitable for cloud computing. It is common that user's input may not be same as the existed keywords such as *google* and *goggle*. If we want to explain with a concrete example, online survey shows that less than 77% of users benefitted by this method within a three months period. The actual traditional method to support keyword search is simple spell check mechanism. This mechanism does not work perfectly for all types of keywords. It is ineffective because it needs more user interaction when the spell check algorithm works , which unnecessarily gives burden to the user. So that user effort is more in this mechanism compare to other. Another reason is sometimes spell check algorithm does not works when the user enters wrong keyword such as toy and boy. And this mechanism cannot differentiate two pre-existed valid words. Thus, the drawbacks of existing system signifies the necessary of new techniques. In this paper, we focus on effective privacy reserving keyword search in Cloud Computing. To the best of our knowledge, we formalize for the first time the problem of effective fuzzy keyword search over encrypted cloud data while maintaining keyword privacy. We use edit distance to compute the keywords similarity and develop wildcard-based technique, for the construction of fuzzy keyword sets. Another efficient technique called gram based technique for constructing fuzzy keyword set is gram based technique. It works based on the grams. Gram of a string can be considered as a substring. Edit operation will affect atmost one character and remaining untouched. Gram is used for constructing inverted list for matching purpose. The above two techniques reduce the need of enumerating all the keywords and reduce the size also. The other technique called symbol based tree traverse search technique is more efficient than above two, in which a multi way tree structure is constructed using symbols for storing the keywords. In this, all trapdoors sharing a common prefix may have common nodes. Thus, the experimental results shows the effectiveness of proposed solution.

The rest of paper consists of four other sections. Section II explains the previous work done on this paper which includes methods on keyword search , the privacy techniques called searchable encryption methods and some other methods. Section III provides the description of proposed system that consists of three modules. Section IV presents the results interms of graphs. And the final section gives conclusion and also future work can be done on this paper.

II. RELATED WORK

KEYWORD SEARCH: The traditional search techniques allows the users to search without using try and see approach for finding relevant information based on the keyword entered by the user. The keyword matching technique includes two types: on-line and off-line. On-line technique searches the data without an index, which is not useful because of its low efficiency. Off-line technique searches the data with an index which is more efficient than on-line. However, these techniques searches the data but fails to achieve the efficiency and privacy.

SEARCHABLE ENCRYPTION: The traditional searchable encryption techniques mostly uses cryptography concept to provide data privacy. In addition to these works, improvements and security definition formulizations are also introduced. The first searchable encryption technique was proposed by Song et al. In this technique, each word in the document is encrypted by using two-layered encryption construction. Goh uses filters to construct the indexes for security. Each filter containing trapdoors of unique words and these are stored in the server. To provide more security and efficiency, Chang et al uses similar index approach, where single encrypted hash table index is built. In this hash table, each keyword associated with the index called trapdoor. However, all these techniques works only exact keyword match occurs. Hence, these are not suitable for cloud computing.

OTHERS: Private matching is the other popular technique which is used to retrieve the matching items secretly. It is mostly used in retrieving data from databases. Sometimes it gives computation complexity.

III. PROPOSED SYSTEM

i)EDIT DISTANCE AND WILD-CARD BASED TECHNIQUE

Edit distance computes the distance between two keywords k1, k2 by using ed(k1,k2). Edit distance is used to measure the keyword similarity. If any dissimilar occurs, three operations can be carried out: a)**Substitution:** substitutes one character with the other character. b)**Insertion:** inserting one character into other character. c)**Deletion:** deleting one character from the other character.

It defines the fuzzy keyword search as follows: Given a collection of datafiles c=(df1,df2,…dfn) are stored in cloud and set of keywords (k1,k2,…kn). When the user sends request this technique searches with edit distance and keyword.

In **wild-card based technique ,** all the variants of keywords to be listed when an operation is performed at the same position. Based on the above approach, we use a wild card to denote the edit operations performed at the same position. This technique edits distance to solve the problems. It includes the following steps:

i)It builds an index with each keyword k. To build index data owner computes f(sk , k).

ii)Construct the secret key sk. This sk is shared among the data owner and user if he is an authorized user.

iii)Searching can be done with secret key sk , keyword k.

iv)Compare the secret key sent by the user and existed key at the data owner. If both are same, returns the requested file.

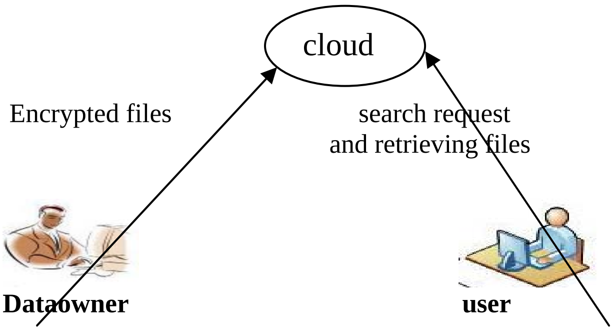
For example, for the keyword FEVER with the pre-set edit distance 1, its wildcard based fuzzy keyword set can be constructed as

FEVER,1 = {FEVER, *FEVER, *EVER, FE*VER, F*VER, FEV*ER, FEV*R, FEVE*R, FEVE*, FEVER*}

ii) GRAM-BASED TECHNIQUE

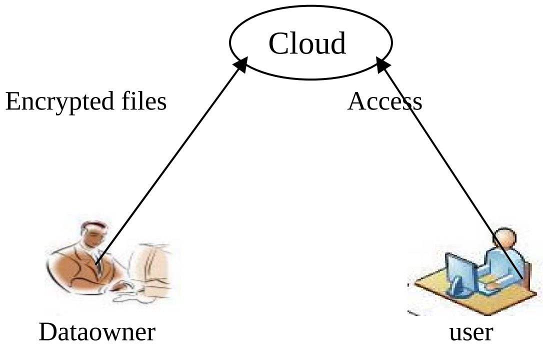
Another efficient technique for constructing fuzzy keyword set is gram based technique. It works based on the grams. Gram of a string can be considered as a substring. Edit operation will affect atmost one character and remaining untouched. In this, gram is used for constructing inverted list for matching purpose. Here, edit operations will affect only one character in the given keyword and all remaining are same.

For example, the gram based fuzzy set EFEVER, 1 for the keyword FEVER can be constructed a {FEVER, EVER,FVER, FEER, FEVR, FEVE}



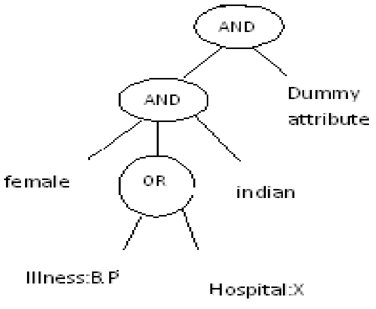
iii) SYMBOL-BASED TREE TRAVERSE SEARCH SCHEME

This technique is more efficient than above two, in which a *multi way tree structure* is constructed for storing the keywords. In this, all trapdoors sharing a common prefix may have common nodes. The keywords in the tree structure can be found by Depth-First Search. The root is associated with an empty set and searching can be done from root to the leaf.



Multi-way tree Structure

The user access structure is used when the user enters more than one keyword. The below example illustrates that, here user wants data about the person who is female, must be an Indian and he/ she either suffers from b.p. or admitted in hospital A.



Multiway tree structure

IV. RESULTS

The concept of this paper is implemented and different results are shown below. The proposed paper is implemented in .net technology on a Pentium-IV PC with 20 GB hard-disk and 256 MB RAM with apache web server.

The propose paper concepts shows efficient results and has been efficiently tested on different Datasets. The Fig 1 describes fuzzy set construction time using distance equal to 1, Fig 2 describes fuzzy set construction time using distance equal to 2 , Fig 3 shows the searching time for edit distance equals to 1 and Fig 4 shows the searching time for edit distance equals to 2 are the real time results compared.

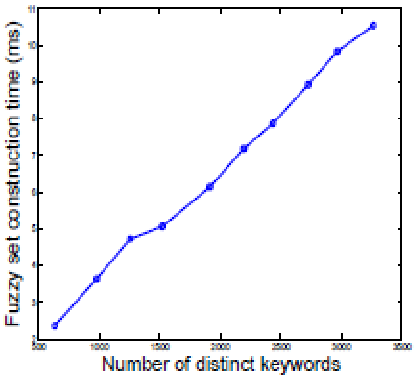


Fig1.fuzzy set construction time using wild card distance=1

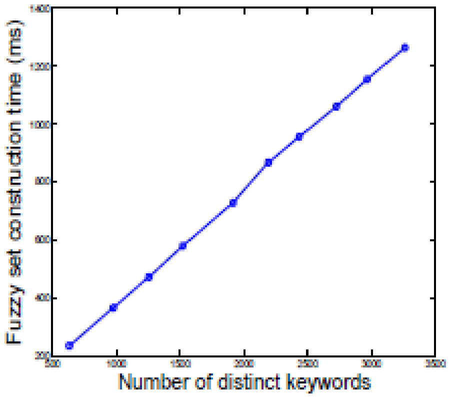
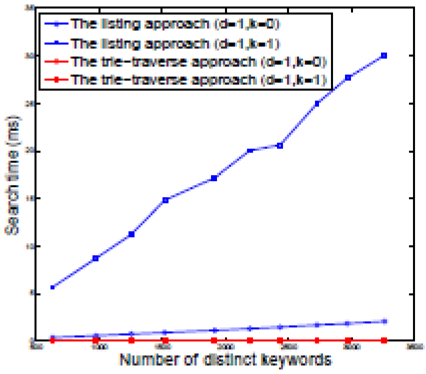
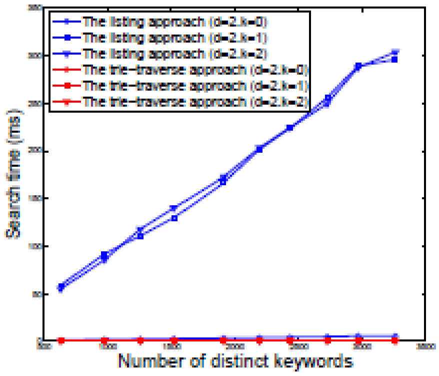


Fig2.fuzzy set construction using wild card distance=2



Searching time for edit distance=1



Searching time for edit distance=2

V .CONCLUSION

In this paper, we try to formalize and solve the problem of providing efficient fuzzy search for remotely stored data in cloud computing.

We design two more advanced techniques (i.e., Gram-based and Symbol-based tree traverse search techniques) to construct efficient fuzzy keyword sets. By providing security, we show that the proposed solution is secure and privacy-preserving. Experimental results demonstrates the efficiency of our proposed solution.

We will continue to research on security mechanisms that support search ranking that sorts the searching results according to the relevance search and semantics that takes into consideration conjunction of keywords, sequence of keywords, and even the complex natural language semantics to produce highly relevant search results.

REFERENCES

1. A. Behm, S. Ji, C. Li, , and J. Lu, "Space-constrained gram-based indexing for efficient approximate string search," in *Proc. of ICDE'09*
2. S. Ji, G. Li, C. Li, and J. Feng, "Efficient interactive fuzzy keyword search," in Proc. of WWW'09, 2009.
3. Google, "Britney spears spelling correction," referenced online at http://www.google.com/jobs/britney.html, July 2009.
4. D. Song, D. Wagner, and A. Perig, "Practical techniques for searches on encrypted data," in Proc. of IEEE Symposium on Security and Privacy'00, 2000
5. M.Belare, A.Boldyreva, and A.O'Neil, "Deterministic and efficiently searchable encryption of rypto 2007, volume 4622 of LNCS, SPRINGER-2007.
6. Y.-C. Chang and M. Mitzenmacher, "Privacy preserving keyword searches on remote encrypted data," in Proc. of ACNS'05, 2005.
7. M. Armbrust and et.al, "Above the clouds: A berkeley view of cloud computing," Tech. Rep., Feb 2009. [Online]. Available at http://www.eecs.berkeley.edu/Pubs/TechRpts/2009/EECS-2009-28.html
8. D. Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in *Proc. of EUROCRYP'04*, 2004.
9. K. N. A. Beimel, P. Carmi and E. Weinreb, "Private approximation of search problems" in proc. Of 38th Annual Symposium on the theory of computing.
10. R. Curtmola, J. A. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: improved definitions and efficient constructions," in Proc. of ACM CCS'06, 2006.