Graduate Research in Engineering and Technology (GRET)

Volume 1 | Issue 2 Article 8

October 2013

USING DIFFERENT SEARCHING SCHEMAS FOR FUZZY KEYWORD SEARCH OVER CLOUD DATA

SYEDA FARHA SHAZMEEN
Department of Information Technology, Balaji Institute of Technology and Science, Warangal, A.P, India, 506331., farhashazmeen@gmail.com

RANGARAJU DEEPIKA
Department of Computer Science and Engineering, Jayamukhi Institute of Technology and Science, Warangal, A.P, India, 506331., rdeepu.515@gmail.com

USING DIFFERENT SEARCHING SCHEMAS FOR FUZZY KEYWORD SEARCH OVER CLOUD DATA

SYEDA FARHA SHAZMEEN [1], RANGARAJU DEEPIKA[2]

[1]Department of Information Technology, Balaji Institute of Technology and Science, Warangal, A.P, India, 506331.
[2]Department of Computer Science and Engineering, Jayamukhi Institute of Technology and Science, Warangal, A.P, India, 506331.
Email: farhashazmeen@gmail.com, rdeepu.515@ gmail.com

Abstract- Cloud Computing is a construct that allows you to access applications that actually reside at a location other than our computer or other internet-connected devices, Cloud computing uses internet and central remote servers to maintain data and applications, the data is stored in off-premises and accessing this data through keyword search. So there comes the importance of encrypted cloud data search Traditional keyword search was based on plaintext keyword search, but for protecting data privacy the sensitive data should be encrypted before outsourcing. Fuzzy keyword search greatly enhances system usability by returning the matching files; Fuzzy technique uses approximate full text search and retrieval. Three different Fuzzy Search Schemas, The wild card method, gram based method and tree traverse search scheme, are dicussed and also the efficiency of these algorithms is analyzed.

Keywords-*Fuzzy keyword, Searching Schemes, Performance issues.*

I. INTRODUCTION

As Cloud Computing becomes prevalent and useful, more and more important and useful information is being centralized into the cloud One of the most popular ways is selectively retrieve files through keyword-based search instead of retrieving all the encrypted files back. The data encryption also demands the preservation of keyword privacy since keywords usually contain important information related to the data files. So in order to improve adaptation of cloud computing [4], first ensure its security.

The data encryption also demands the protection of keyword privacy since keywords usually contain important information related to the data files. The existing searchable encryption techniques do not suit for cloud computing scenario because they support only exact keyword search [2]. This significant drawback of existing schemes signifies the important need for new methods that support searching flexibility. Fuzzy technique uses approximate full text search and retrieval. This means it will match all possible results for a search query despite its form or spelling mistakes/mismatches presence no matter what part of word they will be in. This way it will retrieve "complete" even either your query will be "complte", "compltte" or "compplete". It will show all related results by relevancy and similarity degree.

Since server *S* is a remote server and cannot easily be-trusted, client *C* may need to encrypt the data and store into server *S*. However, client *C* may need to perform a search and retrieve only files that contain a certain keyword. In order to achieve the requirement, client *C* can retrieve all the files that he has saved in server *S*, then decrypt all files and search words by words [1].

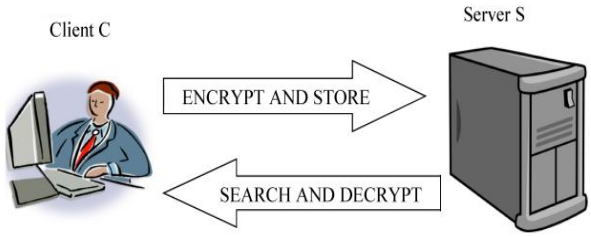


Fig. 1.Searchable Symmetric Encryption

Fuzzy keyword search will return the results by keeping the following two rules.

1. If the user's searching input exactly matches the predefined keyword, then the server is expected to return the files containing that keyword.
2. If there is no exact match or some inconstancies in the searching input, the server will return the closest possible results based on pre-defined similarity semantics.

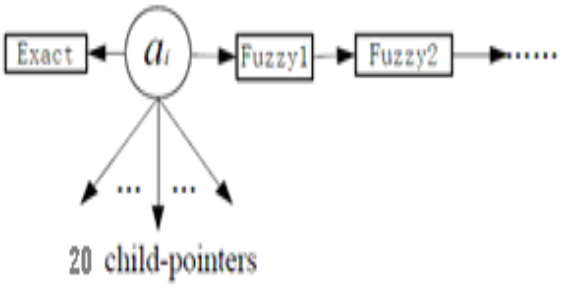


Fig. 2. Fuzzy construction

I(a) Key Word Searching Schemas

The Three Searching Schemas for Fuzzy Keyword Search over Encrypted cloud data are

1. Wildcard – Based Technique

2. Gram - Based Technique
3. Tree Traverse Search Scheme

I(b) Wildcard –Based Technique

This Technique uses fuzzy set edits distance to solve the problems. The Edit distance can be substitution, Deletion and Insertion. By using the Wildcard-Based Technique [9] the Fuzzy keyword search can be more efficient because different edit distance techniques are used which can be used to find for finding the keywords.

I(c) Gram-Based Technique

Another efficient technique for constructing fuzzy set is based on grams [5]. The gram of a string is a **substring** that can be used as a signature for efficient approximate search. While gram has been widely used for constructing
Inverted List for approximate string search, we use gram for the matching purpose.

For example, the gram-based fuzzy set SADVERT, 2 for keyword ADVERT can be constructed as
SADVERT, 2= {ADVERT, AVET, ADVE, ADVR, ADVT}

I(d) Tree Traverse Search Scheme

The search efficiency, tree -traverse search scheme, where a **multi-way tree** is constructed for storing the fuzzy keyword set over a finite symbol set. The key idea behind this construction is that all trapdoors sharing a common prefix may have common nodes. All fuzzy words in the tree can be found by a depth-first search. A conjunctive/sequence of keyword search mechanism will retrieve most efficient and relevant data files [7]; the conjunctive/sequence of keyword search automatically generates ranked results so that the searching flexibility and efficiency will be improved.

A tree-traverse search scheme‟ where a multi-way tree was constructed for storing the fuzzy keyword set and finally retrieving the data. This greatly reduces the storage and representation overheads. It also exploits Edit distance‟ to quantify keywords similarity, to build storage- efficient fuzzy keyword sets to facilitate the searching process [8].
When a user types in multiple keywords. The goal is to efficiently and incrementally compute the records with keywords whose prefixes are similar to those query keywords. We focus on several challenges in this setting. (1) *Intersection of multiple lists of keywords*: Each query keyword (treated as a prefix) has multiple predicted complete keywords , and the union of the lists of these predicted keywords.

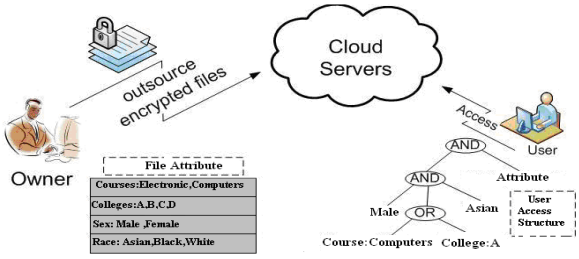


Fig 3. Data Retrieval using Tree Traverse Search Scheme from the Cloud Server

II. KEYWORD SET STORAGE COST AND SEARCHING EFFICIENCY

The different searching schemas are considered, the storage space required and Efficiency in searching fuzzy keyword. We can observe that storage space is decreased and Efficiency in searching the fuzzy keyword over encrypted cloud data is improved [10]. In the present system comparing the storage of keyword set generated for straight forward method, wild card method and gram-based method this is shown in Table 1. If we take into consideration the wild-Card Based Technique edit distance is used for the keyword search the storage space required in high and hence the performance may be decreased. The gram based Technique uses the sub string matching which reduces the storage space but it can be concluded that gram -based method uses less space compared to the other two. But in security aspect wild card method is better. The Tree traverse Search Scheme creates a symbol table for keyword the root is associated with an empty set and the symbols in a trapdoor can be recovered in a search from the root to the leaf that ends the trapdoor. All fuzzy words in the trie can be found by a depth-first search.

Keyword set construction Methods	**Straight forward**	**Wild-card**	**Gram- based**	**Tree Traverse Search Scheme**
Storage space	30 GB	40 MB	10 MB	8 MB
Efficiency In Searching Fuzzy Keyword	Low	Medium	High	High and Efficient

Table (1) Fuzzy keyword search Storage Cost

4(b) characteristics of different Searching Schemas

Table (2) shows the characteristics of the different keyword searching schemas , the below information shows the fuzzy keyword search over encrypted cloud data, where the Privacy and Security are the major concerns in Cloud so the Efficient Techniques are required to efficiently search the keyword.

The characteristics from the straight forward Approach to the Tree Traverse Search Scheme are listed below.

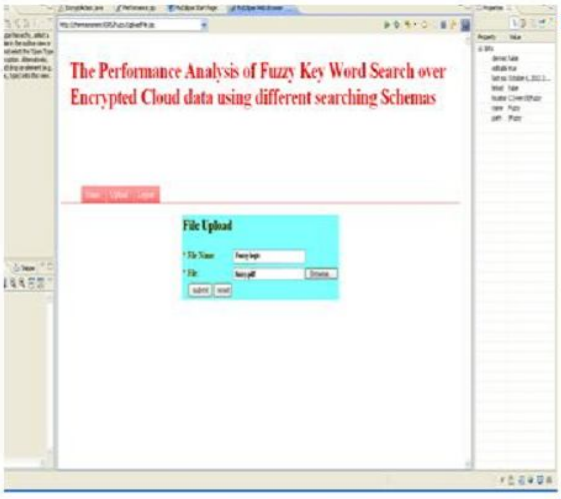
Characteristics	Straight Forward Approach	Wildcard – Based Technique	Gram - Based Technique	Tree Traverse Search Scheme
Hide the Information of the searching Keyword	No	Yes	Yes	Yes
Require a high Perfor--mance machine to compute the algorithm	No	Yes	Yes	Yes
Sequential approach	yes	No	No	No
Fuzzy keyword approach	No	Yes	Yes	Yes
Cloud computing	No	Yes	Yes	Yes
Traverse the Data for the key word search	No	No	No	Yes

Table. 2 A comparison between Wild card, Gram Based, and Tree Traverse Search schemes

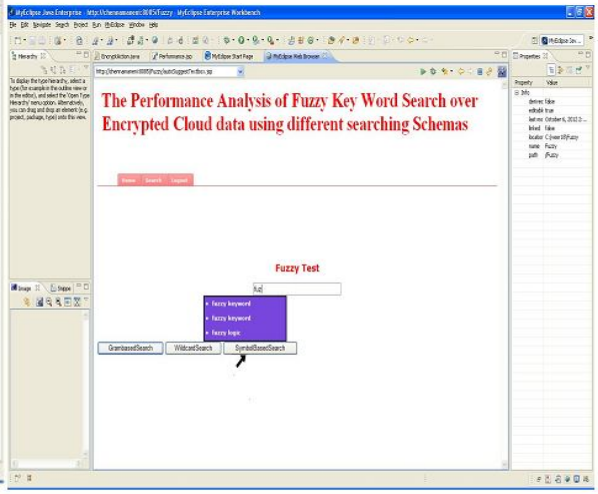
III. IMPLEMENTATION

For the above mentioned techniques for search searching fuzzy keywords, is implemented by using Java ,Java is selected as the programming languages and the other open source API's (Application Programming Interfaces)

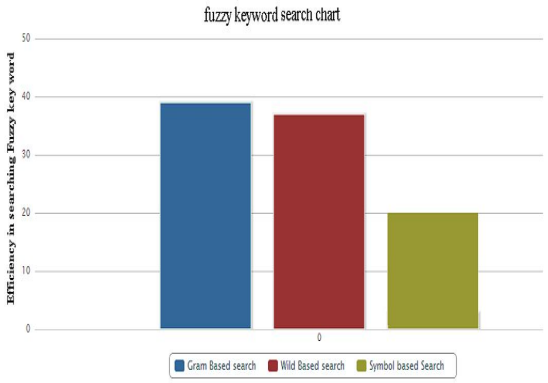
Support the other functionalities. My Eclipse [11] is used as a development IDE (Integrated Development Environment) for Java and library of other technologies are added as external jar (Java Archives) in the eclipse. MyEclipse is built upon the Eclipse platform [12] and integrates both proprietary and open source solutions into the development environment. JFreeChart [13] is an open-source framework for the programming language Java, It is an open source library available for Java that allows users to easily generate graphs and charts. It is particularly effective for when a user needs to regenerate graphs that change on a frequent basis. JFreeChart supports pie charts (2D and 3D), bar charts, line charts, scatter plots, time series charts, and high-low-open close charts.



Fig(4) Shows the File Upload page for a file in Cloud



Fig(5) Shows the different Searching Schemas for Keyword Search



Gram Based search	Wild Based search	Symbol based Search
39	37	20

Fig (6) shows the Efficiency in searching Fuzzy keyword

IV. CONCULSION AND FUTURE WORK

Searching the encrypted data from an encrypted keyword, and then retrieving the data is one of the areas where the security issues occur in the cloud computing scenario. Privacy-preserving fuzzy search for achieving effective utilization of remotely stored encrypted data in Cloud Computing is the recent search technique. The wild card method and gram method to construct fuzzy keyword set and edit distance to quantify keywords similarity are used in this system. Also uses an advanced trie-traverse search scheme, where a multi-way tree is constructed for storing the fuzzy keyword set and finally for retrieving the data. The three searching schemas are taken in to consideration and efficiency of the techniques is measured.

As our ongoing work, we will continue to research on security mechanism that supports for complex natural language semantics to produce highly relevant search result. And multiple semantics like weighted query over encrypted data and checking the integrity of the rank order in the search result.

REFERENCE

[1] Jin Li†, Qian Wang†, Cong Wang†, Ning Cao‡, Kui Ren†, and Wenjing Lou‡,"Fuzzy Keyword Search over Encrypted Data in Cloud Computing"*INFOCOM10.*

[2] Nisha T. M, Kerala Lijo V. P.,"Improving the Efficiency of Data Retrieval in Secure Cloud by Introducing Conjunction of Keywords"**,** *Conference on Advances in Computational Techniques (CACT) 2011 Proceedings published in International Journal of Computer Applications® (IJCA)* 25

[3] Cong Wang, Ning Cao, Jin Li, Kui Ren and Wenjing Lou," Secure Ranked Keyword Search over Encrypted Cloud Data"

[4] Sameer Rajan, Apurva Jairath.Cloud Computing. 2011. The Fifth generation of Computing, International Conference on Communication Systems and Network Technologies.

[5] D. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in *Proc. of IEEE Symposium on Security and Privacy'00*, 2000.

[6] D. Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in *Proc. of EUROCRYP'04*, 2004.

[7] 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.

[8] C. Li, J. Lu, and Y. Lu, "Efficient merging and filtering algorithms for approximate string searches," in Proc. of ICDE'08, 2008.

[9] S.Mishra, S. K. Satapathy, D.Mishra,"Improved search technique using Wildcards or Truncation", In the Proc .of Intelligent Agent & Multi – Agent systems, Published at IEEE Explore 978-1-4244-4711-4/09/, 2009.

[10] Song, D.,Warnger D and Perrig A(2000) "A practical technique for searching on encrypted data" in the proceeding of IEEE Symposium on Security and Privacy ,pp .44-45 May 2000, dio: 10.1109/SECPRI.2000.848445.

[11] MyEclipse, MyEclipse is a commercially available Java IDE and AJAX IDE- http://www.myeclipseide.net/

[12] Eclipse, Open language editor for programming – http://www.eclipse.org/

[13] JFreeChart, JFreeChart is an open source library available for Java that allows users to easily generate graphs and charts http:// www.jfree.org/index.html/