International Research Journal of Computer Science (IRJCS)
Issue 3, Volume 2 (March 2015)
ISSN: 2393-9842
www.irjcs.com

A REVIEW PAPER ON FUZZY SEARCH OVER ENCRYPTED DATA IN CLOUD COMPUTING

Kailash Bhanushali, Neel Gala, Nisha Vanjari
Department of Computer Engineering, KJSIEIT
Ayurvihar Complex, Everard Nagar, Sion, Mumbai 400022
Maharashtra, India

Abstract—***As we know, Cloud Computing is said to be centralized as all the data is stored over cloud. Data maybe or may not be sensitive but then to compromise over privacy is not acceptable. Data must be encrypted before outsourcing. Due to encryption, utilization of data becomes difficult, then also traditional software is capable to retrieve the data even when encryption is performed, but for to get data exact keyword must be matched while search in order to download the required file else you will get empty data. That is no tolerance of even minor typo error and this type of error we found frequently. This leads to low in efficiency and also affects system usability very badly. So here we formalize the problem and able to resolve it by using different techniques even within the limits of keyword and data privacy by achieving them and keeping data encrypt against unauthenticated users and cloud owners. Fuzzy keyword search enhances the system usability by allowing to match the exact or closet match text to the stored keywords and retrieve the approximate closet results. Here we are using edit distance to quantify keywords and using the advanced techniques for construction of keyword set which help to reduce storage and represent overhead. This also help us to ranking search which in turns display most frequented documents on top and later on others with the help of mapping the files according to its use. We are also implementing one of the advanced encryption algorithm for encryption purpose before uploading over cloud to keep it secure even from the cloud users. The more advanced purpose described here is working not only on documents but to on photos and videos with the help of selective encryption algorithm over images and videos.***

Keywords— cloud computing; encryption; fuzzy search; edit distance; ranking search

I. INTRODUCTION

In cloud the data is stored centrally, and there are various types of data stored over the cloud such as social accounts, game data, website login, and many more. The cloud is used because it helps the data owners a relief from storing of data at his place, because storing the data on our own side may be fatal some times; suppose hard-disk failure or any other related problems. The other problem may be as maintenance of data which includes reliability and availability of data and will not getting a high quality service as configuration will not be as good as cloud servers configuration. But cloud too have some drawbacks because cloud servers will not be in trusted domain as of data owners so its user responsibility to encrypt the data before upload. By implementing data encryption, there's overhead of data utilization in effective manner. Moreover, in cloud computing, data owners share their outsourced data with large number of users. Every individual will only retrieve specific data files which they are looking for within a session. To apply this type of system we have to deal with keyword search that retrieve the required files instead of retrieving all the encrypted files.

This keyword search technique allows users too selectively retrieve files of interest and has been widely applied in plaintext search scenarios, such as Google search [1]. Unfortunately, encrypted data restricts user's ability to use the keyword search technique and thus makes the plaintext search methods no use for Cloud Computing.

Besides this, encrypted data files which consist of file name must also be protected as it may also describe the quality and sensitivity information related to the data files. But by encrypting file name the traditional plain text method get totally useless as it only able to search over plain text.

To search over encrypted files, Searchable encryption techniques have been developed in recent years [2]-[10]. Searchable encryption techniques create index of keywords and link that index with the file that related to that keyword [3]-[10]. This technique helps us to search securely and is effective, but not for cloud computing as because all this searchable encryption techniques works only for exact keywords, even a small minor typo error won't allow to retrieve file, and its common that user's input for search might not match the pre-defined keywords such as beautiful instead of beautiful, variations in representation such as R.I.P and RIP, also may be due to not having the detailed knowledge regarding data. So we are implementing the fuzzy keyword search for naïve users by checking spells with the help of different algorithms.

The algorithm implemented traditionally may also not able to solve the problem fully because if some other valid keyword typed by mistake, suppose let's take an example, search for "hate" and mistakenly typed "late", for this type of problem the algorithm won't work, as the word is valid so differentiating among them would be difficult for algorithm. Thus, the drawback leads to implement rather a different algorithm which help us support searching in terms of flexibility, typo errors.

In this paper, we are implementing on fuzzy keyword search over cloud keeping in mind the privacy of document from unauthorized users by applying different encryption algorithm. Fuzzy keyword search help us to enhance working over our system by improving its usability while searching for files by matching the input search text with the pre-defined keywords and will also check for nearest possible match by manipulating users search text with possible values when exact search failed to provide the desired result. We are taking the help of edit distance technique to quantify keywords similarity by implementing the advanced algorithm technique for storing, matching and searching fuzzy keyword sets [Xin Zhou 2006, J. Li 2009]. This algorithms eliminate to need for storing all fuzzy keywords to improve efficiency in terms of privacy as well as overhead of storing large number of keywords by reducing the number of keywords which helps us to retrieve fast data and overhead of matching to all fuzzy keyword is reduced. We are going to implement AES encryption algorithm in order to make secure our documents even from cloud owners by encrypting the documents with the help of AES algorithm and to encrypt image and video we will be using selective algorithm. After analysing, we are able to tell that the solution provided below is comparatively secure as before with increasing the flexibility usage of the system over cloud.

II. LITERATURE SURVEY

Plain text fuzzy keyword search: Currently much of the importance is given to fuzzy search for plain text with the help of fuzzy keyword search by many communities [1]. They were able to solve this problem by rejecting from using try-and-see approach for to search the related information with the help of similar string matching algorithm. At time, if we implemented this string matching algorithm then also it's of no use in terms of privacy because hackers may apply dictionary attack or statistics attack, hence, unauthorized person may be able to get access to the files.

Searchable encryption: Traditional searchable encryption [2]–[8], [10] has been widely studied in the context of cryptography. Most of the algorithm look out for improvements in terms of efficiency and security. The first construction of searchable encryption was proposed by Song et al [3], in which each word in the document is encrypted independently under a special two-layered encryption construction. Goh [4] stated to use Bloom filters to build the indexes for the data files. To get more efficient in terms of search, Chang et al [7] and Curtmola et al [8] both stated similar "index" approaches, in which an encrypted hash table index is created for whole collection. This algorithm consists of index files where each entry of keyword is linked with encrypted set of files which consist of related keywords. . As a complementary approach, Boneh et al [10] presented a public-key based searchable encryption scheme, with an analogous scenario to that of [3]. All these schemes only works with exact search, hence useless for cloud computing.

Complete Search: Bast et al proposed techniques to support "Complete Search," in which a user types in keywords letter by letter, and the system finds records that include these keywords (possibly at different places) [10].

Selective Encryption: Traditionally to encrypt image content, the whole image content is compressed then after compression whole image is encrypted using a standard cipher techniques such as RLE, AES, etc. Normally, encrypted data needs higher transmission and usually not all have that higher bandwidth which makes all this encryption technique not efficient as it must be. Compressed the whole image and later on encrypting it, said to be as fully layered. The other problem with it is all the users may not have same codec functionalities which limits the use of the algorithm over internet. Selecting encryption can be said to as emerging technique, in this only the selected part of image will be encrypted not the whole image. This technique mainly focus to reduce the transfer to encrypted data, while keeping in mind not to compromise with security. This technique also help us to preserve some of the needed codec functionalities. It also very useful in high degree delivery of data and many other positive effects.

III. SOLUTION METHODOLOGIES

System Model:-

In this paper, the cloud system is been associated with the admin, user and cloud server .

The user searches the uploaded data files in cloud system, for which the cloud only permits this service strictly to authorized user, with the help of pre-defined set of different keywords over the encrypted data.

The responsibility of the cloud server is to fetch the correct file on the user interests through identification of file and certain link to a clump of keywords [1]. There are certain protocols in which the implementation of fuzzy keyword scheme generates the result:

1. It is the responsibility of the server to give the correct file on the input of the exact file name given by the user.
2. If the user misses the exact keyword, the server should be able to transmit the nearest result.

Threat Model:-

For the consideration, if the data files are been stored in less secured server, it is possibly that the confidential data may leak through the user request over the cloud. To eliminate this risk factor, the process is been carried through secured manner, which does not affect sensitive information even though the cloud server get the information through the user inputted keyword [2].

Design Goals:-

In this paper,[5] the major obstacle is to provide efficient fuzzy keyword search over encrypted data in cloud yet the confidentiality is preserved. The goals which are been taken mainly into consideration are : I) The storage of fuzzy keyword should be efficient; ii)The design should be efficient on fuzzy keyword; iii) Maintaining the security on the implied scheme.

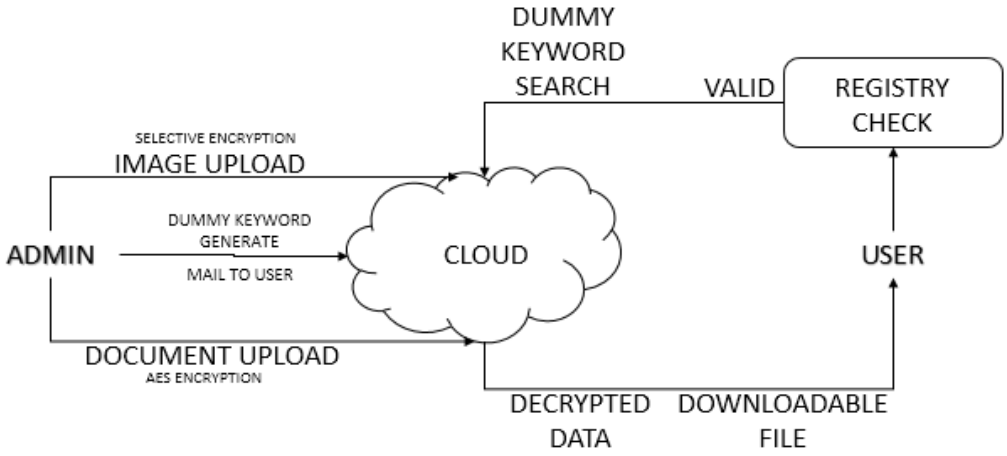

Fig 1: Architecture of Proposed System

Preliminaries:-

Edit Distance Through the detailed study we have chosen the edit distance out of the other quantitatively measure for matching string. There required various operation between respective words (k1, k2) to transform one to another, the operations are Substitution, Deletion and Insertion. Let us assume a keyword *k*, let (W*k*,*c)* denote the set of words k_ satisfying c(k,*k*_) ≤ c for an arbitrary integer *c*.

Advance Techniques for Constructing Fuzzy Keyword Sets:

To enhance straightforward approach and the efficient use of fuzzy keyword practically, with the means of storage and search, we now advance to its advance techniques.[7] The concern over the search matching scheme will be maintained while suppressing the fuzzy keyword. The generality of the search over the encrypted data will be also maintained by focusing on the state of edit distance where d = 1. The reasoning will be same for the increased values of d.

Wildcard-based Fuzzy Set Construct:

As earlier in the straightforward approach we seen that all the keywords operated outlaid to the same position if the list is even. For denoting the edit operations at the same position we proposed to use wildcard-based Fuzzy Set.

Gram Based Fuzzy Search A gram of a string is a substring that can be used as a signature for efficient search. These algorithms answer a fuzzy query on a collection of strings using the following observation: if a string r in the collection is similar to the query string, then should share a certain number of common grams with the query string. This “count filter” can be used to construct gram inverted lists for string ids to support efficient search.

The Symbol-based Trie-Traverse Search Scheme To enhance the search efficiency, we now propose a symbol-based trie-traverse search scheme, where a multiway tree is constructed for storing the fuzzy keyword set {S wi , d } i wi∈W over a finite symbol set.[1]-[3] The key idea behind this construction is that all trapdoors sharing a common prefix may have common nodes. 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.

IV. SUMMARY

In this paper, we able to design fuzzy search over cloud and also eliminate the problem of exact match search, increase the possibility of search by allowing typo error up to some extent. This was possible by implementing two of the advanced technique (i.e. wildcard-based and gram-based technique), they also provide with other much essential benefits such as storage efficient. We also used another efficient search technique (i.e. symbol-based trie traverse scheme) which help us to achieve multiway tree structure with the help of symbols over fuzzy keyword sets. Performing analysis over our system we are able to determine that proposed system is efficient without compromising security issues.

REFERENCES

[1] Fuzzy keyword search over encrypted data in cloud computing" by C. Anuradha.
[2] Implementation of Fuzzy keyword search over encrypted data in cloud computing" by D. VASUMATHI.
[3] Fuzzy keyword search over encrypted data in cloud computing", Illinois Institute of Technology, ISSN: 2321-8134.
[4] Practical techniques for searches on encrypted data" by D. Song, A. Perrig. In IEEE, 2000.
[5] Privacy preserving keyword searches on remote encrypted data" by Y. C. Chang in ACNS, 2005.
[6] Overview on selective encryption of image and video" by A Massoudi in EURASIP, 2008.
[7] Efficient interactive fuzzy keyword search "by J. Feng, G. Li in WWW, 2009.
[8] International Journal of Advanced Research in Computer Science and Software Engineering" Research Paper by P.Kalidas, R.Chandrasekaran.
[9] A. Behm, S. Ji, C. Li,, and J. Lu, "Space-constrained gram-based indexing for efficient approximate string search," in *Proc. of ICDE'09*.
[10] D. Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in *Proc. of EUROCRYP'04*, 2004.