International Journal of Advance Research in Science and Engineering
Volume No.07, Special Issue No.03, April 2018
www.ijarse.com
IJARSE
ISSN: 2319-8354

FUZZY KEYWORD SEARCH BASED ON WILDCARD, GRAM, TRIE-TREE, VITERBI FOR HIDDEN MARKOV MODEL

Pratibha[1]*, Girish Kurdiya[2]*, Sandesh Kharat[3]*

[1,2,3]Department of Computer Engineering

MIT Academy of Engineering Alandi, Pune (India)

ABSTRACT

The Viterbi algorithm is a dynamic programming algorithm for finding the most likely sequence of hidden states called the Viterbi path, which results in a sequence of observed events. Although traditional searchable encryption schemes allow a user to securely search over encrypted data through keywords and selectively retrieve files of interest, these techniques support only exact keyword search. In this paper, for the first time we formalize and solve the problem of effective fuzzy keyword search over encrypted cloud data while maintaining keyword privacy. Fuzzy keyword search greatly enhances system usability by returning the matching files when users searching inputs exactly match the predefined keywords or the closest possible matching files based on keyword similarity semantics, when exact match fails. In our solution, we exploit edit distance to quantify keywords similarity and develop two advanced techniques on constructing fuzzy keyword sets, which achieve optimized storage and representation overheads. Through rigorous security analysis, we show that our proposed solution is secure and privacy-preserving, while correctly realizing the goal of fuzzy keyword search.

Keywords*: *Convolutional Codes, Diarization, Computational Linguistic, Statistical Parsing, Viterbi Parse

I. INTRODUCTION

In this paper with the advancement in cloud computing[1], cloud servers are widely being used for storing data centrally. This includes various social accounts, game data, website login and more type of data. The cloud services provides relief to user as it reduces storage overheads and risk of losing the data due to hardware failures i.e. it might happen the hard disk of our system or due to malicious activity and we would end up losing all the important data. The other problem may be poor maintenance and low configuration service as compared to cloud configuration services. On the other hand cloud also has some drawbacks because cloud servers cannot be trusted by the data owners so it is the user's responsibility to encrypt the data before upload. By implementing data encryption, there is overhead of data utilization in more efficient manner as the data is secured and cannot be accessed by unauthenticated users. Also, in cloud computing, data owners share their outsourced data with large number of users due to which privacy of the data[1] is not ensured. Thus it is required that every individual should retrieve specific data files which they are looking for within a session. To

apply this type of system we need to deal with keyword search that retrieve the required files instead of retrieving all the encrypted files. In plaintext search scenarios such as Google search, the keyword search technique is used which allows users to selectively retrieve the required files . 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. Apart from this, encrypted data files which consist of file name needs to be protected as it may also describe the quality and sensitivity of information related to the data files. But by encrypting file name the traditional plain text methodology get totally useless as it is only able to search over plain text.

II. RELATED WORK

In this paper, we are implementing fuzzy keyword search over cloud without compromising the privacy of our data. By employing fuzzy keyword search the usability of our system is enhanced. Users can search their text with possible values and get the desired result when exact keyword match fails. This failure of exact keyword could be because of some spelling or morphological error. Thus fuzzy keyword search helps to overcome this and give desired results to the user. In our proposed system, edit distance technique to quantify keywords similarity by implementing the advanced algorithm technique or storing, matching and searching fuzzy keyword sets. These algorithms eliminate the 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 shall be implementing AES encryption algorithm[2] before uploading our documents over the cloud servers. This is done to ensure secure and privacy of our data against unauthenticated user .Fuzzy keyword search would be then implemented using N grams and wildcard-based technique[1].

III. METHODOLOGY

Fuzzy keyword search over cloud without compromising the privacy of our data.

The block architecture is as shown in Fig. 1:

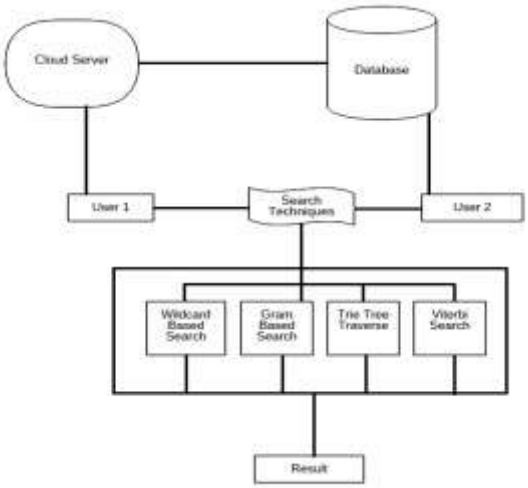

Fig1. General Architecture

The idea of this proposed application is search over cloud without compromising the privacy of our data:

- **SEARCHABLE ENCRYPTION**

Traditional searchable encryption has been of much importance in cryptography. In this approach each word is encrypted independently under two layer encryption construct to provide security. Unfortunately, the scheme is not secure against statistical analysis across multiple queries and can leak the positions of the queried keywords in a document. The searching overhead is linear since each word in the file is encrypted independently. To achieve more efficient search, Goh put forward to use Bloom filters[4] to construct the index for each file an makes this make the search scheme independent of the file encryption. Also, the complexity of each search request is approximately proportional to the number of files in the collection. Curtmola Et Al. proposed the formal security notion of searchable encryption. Furthermore, they put forward similar index approaches, where one encrypted hash table index is constructed for the entire file collection. In the index table, every entry consisted of the trapdoor of a keyword and an encrypted set of related file identifiers. Bao Et Al[3]. also proposed a searchable encryption scheme in multi-user setting, where a group of users can share data in away that can contribute searchable contents and can search an encrypted file collection without disclosing their secrets.

- **COMPLETE SEARCH**

In complete search user types the keyword letter by letter and system retrieves all the records that contain the keyword.

- **CONSTRUCTION OF EFFECTIVE KEYWORD SEARCH**

The key idea behind our secure fuzzy keyword search is two-fold:

1) Building up fuzzy keyword sets that incorporate not only the exact keywords but also the ones differing slightly due to minor typos, format inconsistencies, etc. Download Files View Agent Account View Files.

2) Designing an efficient and secure searching approach for file retrieval based on the resulted fuzzy keyword sets.

- **WILDCARD BASED TECHNIQUE**

In this approach, all the variants of the keywords have to be listed even if an operation is performed at the same position. This method can be used to denote edit operations at the same positions.

- **GRAM BASED TECHNIQUE**

Gram or 1-Gram has been widely used for constructing inverted list for approximate string search. We will use gram for the matching purpose. We propose to utilize the fact that any primitive edit option will affect at most one special character of the keyword, leaving all the remaining characters untouched.

- **SYMBOL-BASED TRIE-TRAVERSE SEARCH SCHEME**

A multi-way tree is constructed for storing the fuzzy keyword set over an infinite symbol set. All fuzzy words in the trie can be searched through Depth- first search. The key idea behind this construction is that all trapdoors sharing a common prefix may have common nodes.

- **VITERBI SEARCH**

Viterbi algorithm or Viterbi search is a dynamic programming algorithm for finding the most likely sequence of hidden states-called the Viterbi path that results in a sequence of observed events.

IV. CONCLUSION

In this paper , we proposed Fuzzy Keyword Search based on mainly these techniques i.e. Wildcard , Gram , Trie-tree and Viterbi Algorithm . It allows a User to securely search over encrypted cloud data while maintaining keyword privacy. One can perform searching over file names and contents both and on very different file formats and file types. Fuzzy keyword search greatly enhances system usability by returning the matching files. The users searching inputs exactly match the predefined keywords or the closest possible matching files based on keyword similarity semantics when exact match fails.

REFERENCES

[1] Flexible Wildcard Searchable Encryption System,Yang Yang, Member, IEEE, Ximeng Liu, Member, IEEE, Robert H.Deng, Fellow, IEEE, Jian Weng. DOI 10.1109/TSC.2017.2714669, IEEE

[2] Modified Viterbi Algorithm for Decoding of Block Codes Zolotarev V.V. and Grinchenko N.N., Ovechkin G.V., Ovechkin P.V. 978-1-5090-6742-8/17/$31.00 2017 IEEE

[3] Fu Z, Wu X, Guan C, et al. Toward efficient multi-keyword fuzzy search over encrypted outsourced data with accuracy improvement [J]. IEEE Transactions on Information Forensics and Security, 2016, 11(12): 2706-2716.

[4] Bloom B H. Space/time trade-offs in hash coding with allowable errors[J]. Communications of the ACM, 1970, 13(7): 422-426.