IMPACT: International Journal of Research in Engineering & Technology (IMPACT: IJRET)
ISSN(P): 2347-4599; ISSN(E): 2321-8843
Vol. 4, Issue 6, Jun 2016, 15-20

FUZZY KEYWORD SEARCH OVER ENCRYPTED DATA IN CLOUD COMPUTING

MOHIT TIWARI, RAKSHIT MAHAJAN, SHRADHA AHUJA, SAHIL RAWAT & VISHRUT MITTAL

Department of Computer Science Engineering, GGSIPU Rohtak Road, East Paschim Vihar Delhi, India

ABSTRACT

With the evolution in Cloud Computing more and more sensitive data is being incorporated into the cloud. To ensure security and privacy these data are first encrypted before being uploaded onto the cloud servers thus making search a complicated task. Although in traditional cloud computing encryption searching schemes allows user to search encrypted data through keywords securely. These techniques employed exact keyword search and will fail if there are any morphological variants or spelling errors. This leads to low in efficiency and also affects system usability very badly. Fuzzy keyword search increases the system usability by allowing matching the exact or closet match text to the stored keywords and retrieving the approximate closest results. We shall be using edit distance to quantify keywords. We ensure the privacy of the data against unauthenticated users by encrypting the data using AES encryption before uploading to the cloud servers. We tend to resolve this problem by using a cloud server and employing fuzzy keyword search based on N grams. Thus efficiency of our proposed system would be demonstrated through experimental results.

KEYWORDS: Encryption, Fuzzy Keyword, Cloud Computing

INTRODUCTION

With the advancement in cloud computing, 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's 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 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 [1].

In plaintext search scenarios such as Google search, the keyword search technique is used which allows users to selectively retrieve the required files[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. 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.

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 for 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 before uploading our documents over the cloud servers. This is done to ensure secure and privacy of our data against unauthenticated users. Fuzzy keyword search would be then implemented using N-grams and wildcard-based technique.

RELATED WORK

Plaintext 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 [2]. This problem was solved by rejecting the idea of using of try and see approach for searching related work and instead using using string matching algorithms. But again this method lacks in terms of privacy as hackers may apply statistical or lexicon attack and gain unauthorized access to the files.

Searchable Encryption

Traditional searchable encryption [3]-[9], [11] 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 [5] put forward to use Bloom filters 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. [9] 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. [11] also proposed a searchable encryption scheme in multi-user setting, where a group of users can share data in a way 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.

PROPOSED SYSTEM

Main Modules

- AES Encryption
- Wildcard-Based Technique
- Grams-Based Technique

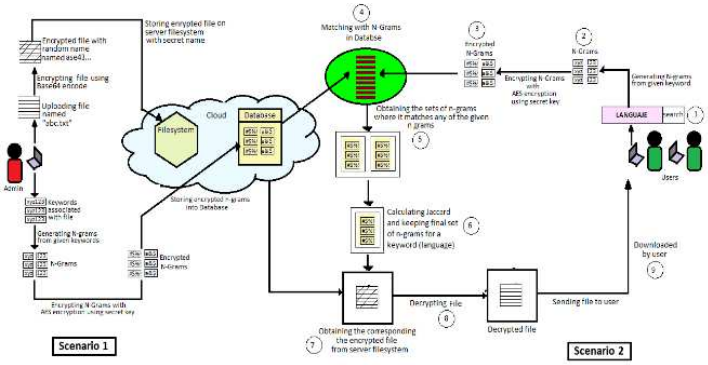

Figure 1: Proposed Architecture of the System

Advanced Encryption Standard (AES)

AES is a block cipher technique with block size of 128 bits. It mainly allows three different key lengths: 128, 192 or 256 bits. We propose to encrypt our data by using AES with 128 bit key length. For 128-bit keys the encryption process consists of 10 rounds. In each case all the rounds are identical other than the last round. A 4 byte word is formed from 16 byte encryption key which is expanded into key schedule consisting of 44 4-byte words. A 4x4 matrix of bytes known as state array is obtained from 128-bit input block. The input state is XORed with first four words of the schedule before any round based processing could start [12].

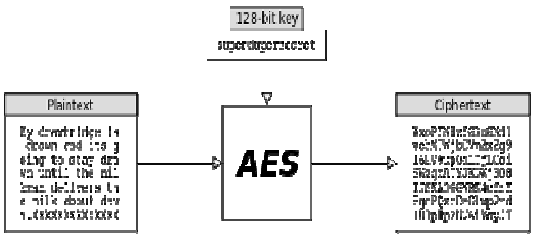

Figure 2

Wildcard – Based Technique

This is straightforward approach where all the variants of the keywords have to be listed even if an operation is performed at the same position. Based on the above observation, we use wildcard to denote edit operations at the same position. The wildcard-based fuzzy set edits distance to solve the problems of searching encrypted files.

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

SCASTLE, 1 = {CASTLE, *CASTLE,*ASTLE, C*ASTLE, C*STLE, CASTL*E, CASTL*, CASTLE*}.

Edit Distance

- Substitution
- Deletion
- Insertion
- **Substitution:** replacing one character by another in a word;
- **Deletion:** deleting a single character from a word;
- **Insertion**: inserting a single character into a word [1].

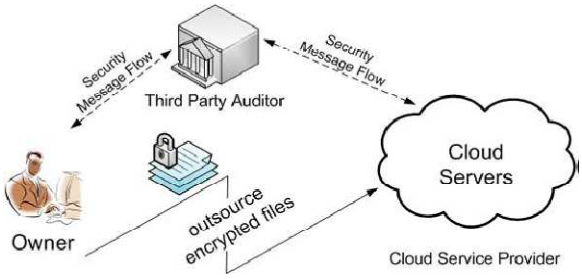

Figure 3

Gram – Based Technique

One of the most efficient technique for constructing fuzzy set is based on grams. The gram is a substring of a string that is used for making approximate search. While gram has been widely used for constructing inverted list for approximate string search, we use gram for the matching purpose. We shall utilize the fact that any primitive edit operation will affect at most one specific character of the keyword, leaving all the remaining characters unused. In other words, the relative order of the remaining characters after the primitive operations is always kept the same as it was before the operation.

For example, the gram-based fuzzy set SCASTLE, 1 for keyword CASTLE can be constructed as

{CASTLE, CSTLE, CATLE, CASLE, CASTE, CASTL, ASTLE}[1]

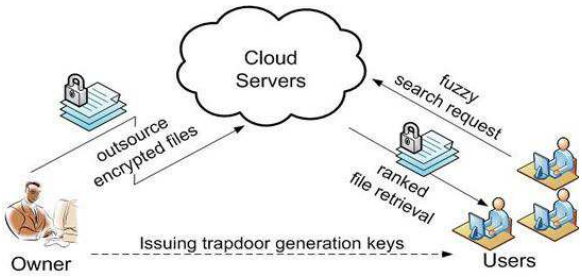

Figure 4

CONCLUSIONS

In this paper, we have solved the problem secure and privacy preserving fuzzy search for effective utilization of encrypted data uploaded onto the cloud servers. We have created fuzzy sets using gram and edit distance technique. After thorough security analysis we have found that our proposed system is secure and privacy preserving while realizing the goal of Fuzzy keyword search.

REFERENCES

1. A REVIEW PAPER ON FUZZY SEARCH OVER ENCRYPTED DATA IN CLOUD COMPUTING by Neel Gala ISSN:2393-9842
2. *Fuzzy keyword search over encrypted data in cloud computing*" by T. Balamuralikrishna.
3. *Implementation of Fuzzy keyword search over encrypted data in cloud computing*" by D. VASUMATHI.
4. *Fuzzy keyword search over encrypted data in cloud computing*", Illinois Institute of Technology, ISSN: 2321-8134.
5. *Practical techniques for searches on encrypted data*" by D. Song, A. Perrig. In IEEE, 2000.
6. *Privacy preserving keyword searches on remote encrypted data"* by Y. C. Chang in ACNS, 2005.
7. Overview on selective encryption of image and video" by A Massoudi in EURASIP, 2008.
8. *Efficient interactive fuzzy keyword search* "by J. Feng, G. Li in WWW, 2009.
9. International Journal of Advanced Research in Computer Science and Software Engineering" Research Paper by P. Kalidas, R. Chandrasekaran.
10. A Behm, S. Ji, C. Li,, and J. Lu, "*Space-constrained gram-based indexing for efficient approximate string search*," in *Proc. of ICDE'09*.
11. Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, "*Public key encryption with keyword search*," in *Proc. of EUROCRYP'04*, 2004.
12. Implementation of Fuzzy Keyword Search Over Encrypted Data in Cloud Computing by ChandniChandawalla.