International Journal of Emerging Trends in Engineering and Basic Sciences (IJEEBS)
ISSN (Online) 2349-6967
Volume 3, Issue 2 (March-April 2016), PP.10-15

Fuzzy Keyword Search Encrypted Data On Cloud Computing

Pradnya S. Borkar
(Department of Computer Science & Engg, Priyadarshini J.L.College of Engineering, Nagpur/ RTMNU, India)

Abstract:- In this paper the application developed for anticrime branch based on cloud is explained, which deals with criminals data to be uploaded into the encrypted format. This project is for anti-crime branch, where registered officials create and maintain confidential data files of crimes and criminal's details throughout the world. Lawyers can register themselves to read the files in decrypted mode related to crimes and criminals, as a subject case study of their degree. If there is any crime, anti-crime branch search the relevant crime on this cloud in the past so that they can detect the nearby criminal. This application is developed by using PHP as a sever side scripting language with Dreamweaver. The AES (Advanced Encryption Standard) cryptography algorithm is used for storing data confidentially so that no one can read it. The Fuzzy search concept is used to autocorrect the misspelled word automatically by using INSERTION, DELETION and SUBSTITUTION rule.

Keywords: - fuzzy keyword encrypted data, cloud.

I. INTRODUCTION

This is a cloud based application which deals with crimes and criminals data uploaded on cloud in the encrypted form. This project is for anticrime branch; therefore the authorized person's of anticrime branch create and maintain the criminals and crimes record. Encryption and decryption process is used for security purpose. For maintaining security AES algorithm is used, and fuzzy search concept is used for searching crimes & criminals data. As Cloud Computing becomes prevalent, more and more sensitive information are being centralized into the cloud. For the protection of data privacy, sensitive data usually have to be encrypted before outsourcing, which makes effective data utilization a very challenging task. 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. That is, there is no tolerance of minor types and format inconsistencies, which on the other hand typical user searching behavior and happen very frequently. This significant drawback makes existing techniques unsuitable in Cloud Computing as it greatly affects system usability, rendering user searching experiences very frustrating and system efficiency very low. In this application i tried to solve the problem of effective fuzzy keyword search over encrypted cloud data while maintaining data 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. The main concept to be used in the application is cloud. Cloud Computing is a general term for anything that involves delivering hosted services over the internet. The data is in encrypted form on cloud to provide security .The user's are given a search pages ,where they can search any data using fuzzy logic that makes searching easier. The downloaded data is decrypted and provided to lawyer in readable format. These services are broadly divided into three categories:

- **IaaS** - Infrastructure as a Service
- **PaaS** - Platform as a Service
- **SaaS** - Software as a Service

Infrastructure-as-a-Service (IaaS)

Cloud Providers offers Infrastructure as a Service throughout data-centre space, and servers, as well as network equipment such as routers/switches and software for businesses. These data-centres are fully outsourced, need

not lift a finger, upgrade an IOS or re-route data. Although this is the base layer, it allows for scalability and reliability as well as better security than an organization may have in a local or local data centre.

Platform-as-a-Service (PaaS)

Provisioning a full hardware architecture and software framework to allow applications to run is the essence of Platform-as-a-Service.

Software-as-a-Service (SaaS)

Software-as-a-Service is the process of provisioning commercially available software but giving access over the net.

II. FUZZY KEYWORD SEARCH

To ensure security during information retrieval, a searchable encryption mechanism is employed. In a standard searchable encryption scheme supporting an exact keyword match is inconsistent with a casual user search behaviours. Normal user search queries will have the types and representation irregularities which may not match the pre-set keyword strings. A user searching for 'APPLE' can accidently type 'APLE' and another person may query for 'PO BOX' instead of 'P.O.BOX' because he is ignorant about the stored keywords.
Thus, the main focus is on enabling effective privacy preserving fuzzy keywords search for information stored in cloud environment. Fuzzy keyword search arguments system usability by returning the matching files when users searching inputs exactly match the predefined keywords or the closes possible matching files based on keyword similarity semantics, when exact match face. Edit distance is used to quantify keywords similarity and for the development of a novel technique, i.e., wildcard-based technique, for constructing fuzzy keyword sets. This technique eliminates the need for counting all the fuzzy keywords and total size of the fuzzy keyword sets is significantly decreases.

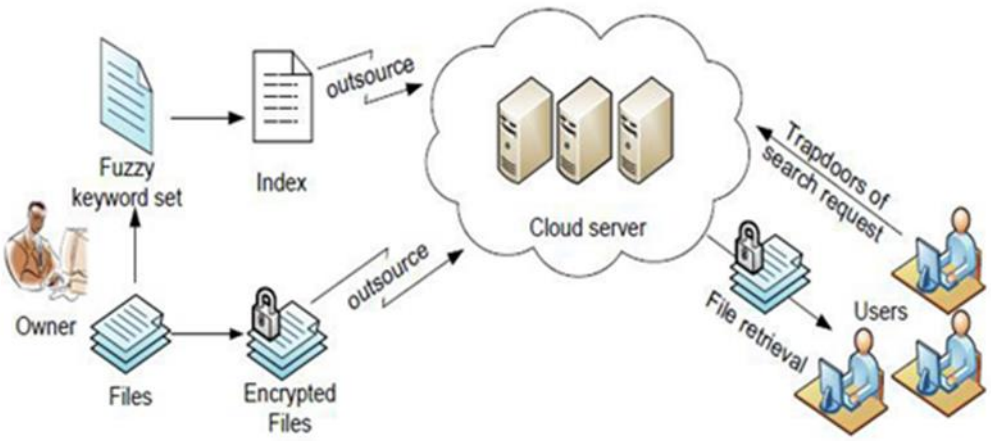

Fig 1: Architecture of fuzzy keyword search

III. LITERATURE SURVEY

In [1]the author proposed a unique data structure called Secure File Object (SFO) to enable keyword search over encrypted cloud data .When a data owner wants to upload a data file to the cloud storage, the client side application will create and attach a SFO to the encrypted data file before uploading to the cloud storage. Each SFO contains information that describes the uploading data file. During SFO creation, the client-side application extracts unique keyword from the uploading data file and encrypts them to create a list of encrypted keywords that will be stored in the SFO. When a user wants to search for a specific keyword, the user will submit the keyword to the data owner and the data owner will compute the search capability by encrypting the keyword with the same key that used to generate the list of encrypted keywords in the SFO. The user can submit the returned search capability from the data owner to the cloud server. The cloud server will return the encrypted data file if the list of encrypted keywords in the SFO contains the search capability. This proposed scheme with Semantic Search over Encrypted Data in Cloud Computing SFO is implemented by the authors to provide simple keyword search over encrypted cloud data. The major drawback of the SFO scheme is that the scheme only supports keyword search using the exact keyword as it appears in the data file. If there are any types of the keywords that used to generate the search capability, the cloud server will fail to locate the correct encrypted data file.

In [2]the "Wildcard-based Fuzzy Set Construction (WFSC)" scheme is proposed which is used fuzzy keyword search over encrypted cloud data. The key concept behind WFSC is maintaining an index that covers all possible variations of a keyword within a predefine edit distance. Instead of simply encrypting the keywords extracted from the data file, WFSC expands each extracted keyword into a set of modified keywords by inserting wildcard character into the keyword. The number of wildcard character used to modify the keyword is based on a predefine edit distance value. Table 1 shows the modified keyword set of the keyword "student" using the wildcard character '*' when the predefine edit distance value equals to 1. Each modified keywords in the set will be hashed with a secured hash function to create a trapdoor. The trapdoor will be appended by the encrypted information that describes the uploading data files that contain the keyword and the original keyword to form an index entry. The collection of index entries will form an index file and it will be uploaded to the cloud storage along with all the encrypted data files that addressed by the index file.

Table 1: Modified keyword set of keyword "student" when the predefine edit distance value equals to 1:

Edit Distance	**Modified Keyword Set**
1.	*student, *tudent, s*tudent, s*udent, st*dent, st*udent, stu*dent, stu*ent, stud*ent, stud*nt, stude*nt, stude*t, studen*, studen*t, student, student*

When a user needs to search for a specific keyword, the user will inject the keyword with wildcard character to compute the modified keyword set based on the predefine edit distance value and hash each modified keywords to create the trapdoor set. The trapdoor set will be submitted to the cloud server and the cloud server will search the index file and compare the trapdoor in each index entry with each trapdoor in the received trapdoor set. The cloud server will return Semantic Search over Encrypted Data in Cloud Computing the matched encrypted index entries to the user and the user can decrypt the index entry to retrieve the information of the data files that contain the keyword.

There are several weaknesses in WFSC if the user tries to expand the search coverage by increasing the predefine edit distance value. The increase in the edit distance value will cause a huge increase in the size of the modified keyword set. Table 2 shows the modified keyword set of the keyword "student" using the wildcard character '*' when the predefine edit distance value equals to 2. Our examples of keyword "student" show the size of the modified keyword set is 16 when the edit distance value equals to 1 and the size of the modified keyword set has increased hugely to 122 by simply increase the edit distance value to 2. The size of the index file will increase rapidly by increasing the edit distance value and the search performance will be degraded due to the increases in the index size. Furthermore, the quality of the search will also degraded due to more unrelated keywords can be returned in the result.

Table 2: Modified keyword set of keyword "student" when the predefine edit distance value equals to 2

Edit Distance	Modified Keyword Set
2.	**student, **tudent, **udent, *s*tudent, *s*udent, *st*dent,*st*udent, *stu*dent, *stu*ent, *stud*ent, *stud*nt,*stude*nt, *stude*t, *studen*,*studen*t, *student,*student*, *t*dent, *t*udent, *tu*dent, *tu*ent, *tud*ent,*tud*nt, *tude*nt, *tude*t, *tuden*, *tuden*t, *tudent,*tudent*, s**dent, s**tudent, s**udent,s*t*dent,s, *t*udent, s*tu*dent, s*tu*ent, s*tud*ent, s*tud*nt, s*tude*nt,s*tude*t, s*tuden*, s*tuden*t, s*tudent, s*tudent*, s*u*dent,s*u*ent, s*ud*ent, s*ud*nt, s*ude*nt, s*ude*t, s*uden*,s*uden*t, s*udent, s*udent*, st**dent, st**ent, st**udent, student*, student**

Instead of injecting keyword with the wildcard character, DFSC uses a dictionary to pull in only the valid words that are within the range of the predefine edit distance to form the modified keyword set. The authors showed the size of the index file created by DFSC is much smaller than WFSC when the predefine edit distance increased. Even DFSC shows better storage usage than WFSC, DFSC inherited the search quality degradation problem because more unrelated keywords can still be pulled into the modified key set from the dictionary as edit distance value increases. Furthermore, DFSC does not support variations of newly invented words or keywords that contain multiple typos because they are words that cannot be found in the dictionary.
The following technical areas are used in the application cryptography:

1) Cryptography
2) Encryption and Decryption using AES algorithm
3) Fuzzy keyword search

Cryptography is the technique used for the secure data from unauthorized users, it means the plaintext message convert into ciphertext message. For converting plaintext to ciphertext and cipher to plaintext message using the

encryption and decryption key. Encryption means the plaintext message is converted into ciphertext message and Decryption means the cipher text is converted into plaintext.

As security on cloud is very important, we have used AES algorithm to achieve file security so as to make data non-readable. However, data encryption makes effective data utilization a very challenging task given that there could be a large amount of outsourced data files**.** Moreover, in Cloud Computing, data owners may share their outsourced data with a large number of users. The individual users might want to only retrieve certain specific data files they are interested in during a given session.

Fuzzy search:

We have used fuzzy search technique to search keyword from encrypted data on cloud. Which is also used to autocorrect misspelled words. In This project we use the levenshtein algorithm for fuzzy search.
Levenshtein Algorithm:

Step1: In this algorithm, first we take input.

($input = $_POST[“t1”];)

No shortest distance found i.e. no match is found give the output is-1($shortest = -1);

Step2: If the input is not match, then it will checked the another step using loop for each($arr as $word){

Calculate the distance between the input word and current word.

$lev = levenshtein($input, $word);

Step3: In this step, if the current input word are not matched then it will give the output as

Zero(0) it means that input word are exactly matched go further the break loop.

If($lev = =0)

$closest = $word;

$shortest = 0;

Break;}

Step4: If the word is nearby matched then distance is less than shortest.

if($lev <= $shortest || $shortest<0)

set closest matched and shortest distance.

$closest = $word;

$shortest = $lev;

This system includes a user application, a cloud computing server, and a cloud storage server. The user application is hosted and maintained by the user party and is considered as a trusted component while the cloud computing server and the cloud storage server is hosted and maintained by a third party and are considered as non-trusted component. Any communication channel between the user application and the cloud servers also considered as non-trusted because it can be targeted by other attackers. This application is acting as the interface to handle all communications between the user and the cloud computing servers.

REFERENCES

1. C.Bagyalakshmi, Dr.R.Manicka Chezian "A secured searching in cloud data using cryptographic technique" international journal of advanced research in computer engineering & technology (ijarcet) volume 1, issue 7, september 2012.

2. Sandesh Molane, Sachin Mate, Shubham Mahamine, Rishikesh Nikam & U. A. Mande "Faster Retrieval of Data from Cloud using Feature Extraction" International Journal of Interdisciplinary Research (IJIR) Vol-2, Issue-5, 2016 ISSN: 2454-1362.

3. Pooja Malwar, Shruti Gedam, Priya Turankar ,"Fuzzy Keyword Search Over Encrypted Data In Cloud Computing" International Journal For Engineering Applications And Technology , ISSN 2321 – 8334 March 2013.

4. Aarti Singh, Manisha Malhotra, "Security concerns at various level of cloude computing paradigm: A Review" International Journal of computer network and application Volume 2,Issue 2, March-April (2015).

5. P. Kalidas , R. Chandrasekaran , Dusmant Kumar Sahu , G.Michale "Efficient Interactive Fuzzy Keyword Search Over Encrypted Data in Cloud Computing " International Journal of Advanced Research in Computer Science and Software Engineering Volume 3, Issue 4, April 2013.

6. Atul Kahate "Cryptography and Network Security" Second Edition.

7. http://www.calvin.edu/-pribeiro/otherlnks/fuzzy/home.html.