ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 4, April 2017

# Fuzzy Keyword Search over Encrypted Data

Utkarsh Joshi[1], Neeraj Vishwakarma[2], A.Murugan[3]

U.G. Student, Department of Computer Science and Engineering, SRM University, Chennai, India[1, 2]

Associate Professor, Department of Computer Engineering, SRM University, Chennai, India[3]

**ABSTRACT**:In today's world utilization of resources effectively is very necessary. With the increased rate of growth and adaptation of cloud computing, daily more and more sensitive information is being centralized onto the cloud and to protect sensitive information, data must be encrypted. In this paper, we focus on secure storage using Advanced Encryption Standard (AES) and information retrieval by performing fuzzy keyword search on this encrypted data. Traditional searchable encryption schemes provide a range of approaches to search on encrypted data, but they only support exact keyword search. Exact keyword search is not suitable for cloud storage systems, because it doesn't allow users making any spelling errors or format inconsistencies, which greatly reduce the system usability. To the best of our knowledge, the relatively most feasible scheme published so far which supports fuzzy keyword search is the "Wildcard-based Fuzzy Set Construction". The Wildcard Based technique returns the matching files when users searching inputs exactly match the predefined keywords or the closest possible matching files based on similarity keyword semantics, when exact match fails..

**KEYWORDS**:AES Encryption, Jaccard, Wild-card technique, codeigniter PHP Framework, N-Grams, Cipher text..

## I. INTRODUCTION

With the rapid development of cloud computing, there is a growing trend of cloud storage usage. Since the data is encrypted, the searching of documents which contain specific keywords becomes rather difficult. We may come up with a solution that the user downloads all the encrypted data and decrypts them with his/her secret key, and then he/she uses normal search methods to search documents containing specific keywords. This approach is obviously inefficient and requires users to have strong storage capacity. According to searchable encryption solutions can be divided into three classes:

1. Index-based approach
2. Search in the encrypted data
3. Using secret sharing

The key feature of AES Encryption is to produce complex cipher text which is very difficult to decrypt without a secret key.

## II. RELATED WORK

Currently much of the importance is given to fuzzy search for plain text with the help of fuzzy keyword search by many communities. This problem was solved by rejecting the idea of using of try and see approach for searching related work and instead using string matching algorithms.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 4, April 2017

Main modules:

- Base64: - This is an encryption technique which encrypts the binary data. In our implementation we have encoded file name using this technique. First it converts the text to binary form then it encodes this into cipher text.

- N-Grams: - It is contiguous sequence of n items. An N-Gram of size 1 referred to as “Unigram”, size 2 is a “bigram” and size 3 is a “Trigram”. In our project we have used Trigrams to brake the keywords.

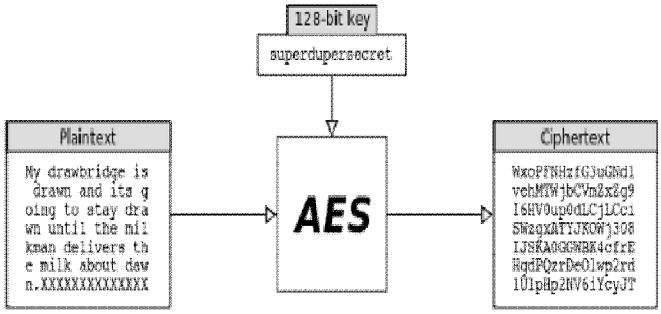
- AES Encryption: - This encryption is used to encrypt the Trigrams with the help of secret key.

  a) Text file is used as input in AES encryption technique which produces the cipher text.
  b) In our implementation we have used 256 bit key size which is used to perform 14 round of permutation and substitution (p-box and s-box) operation.
  c) Each round comprise of four sub-processes. The first round process is:

     1) Byte Substitution
     2) Shift Rows
     3) Mix Columns
     4) Add round key

- Decryption Process: – The process of decryption of an AES cipher text is similar to the encryption process in the reverse order. Each round consists of the four processes conducted in the reverse order

  1) Add round key
  2) Mix columns
  3) Shift rows
  4) Byte substitution



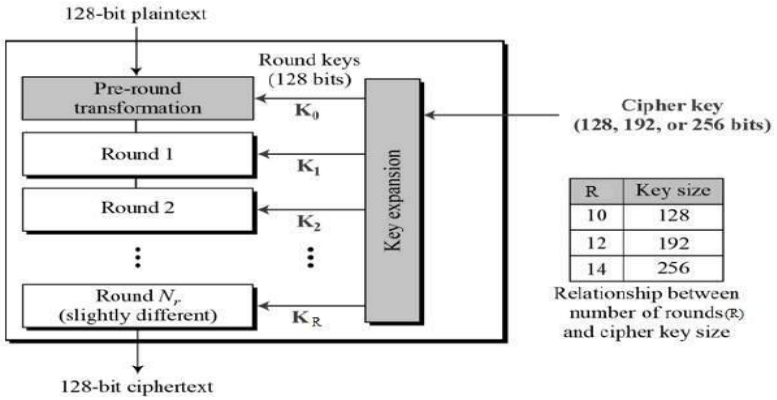
ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 4, April 2017




| R | Key size |
|---|---|
| 10 | 128 |
| 12 | 192 |
| 14 | 256 |

Relationship between number of rounds(R) and cipher key size

Fig 1: AES Encryption

## III. LITERATURE SURVEY

**Plain Text Fuzzy Keyword Search:**
The most importance given in today's era is to fuzzy search for plain text with the help of fuzzy keyword search. The problem was solved by rejecting the use of try-and-see approach to search the relevant information with the help of similar matching string algorithm. If we apply this string matching algorithm then also it's of no use considering the privacy of the data 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 are widely studied in the cryptography.
Most of the algorithm look out for improvements in terms of efficiency and security. In the first construction of searchable encryption each word in the document is encrypted independently under a special two layered encryption construction. Bloom filters were used to build the indexes for the data files. To get more efficient in terms of search 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.All these schemes only works with exact search, hence useless.

**Complete Search:**
In this the user types in key-words letter by letter, and the system finds records that include these keywords (possibly at different places).

## IV. SYSTEM ARTCHITECTURE

Architectural diagram shows the relationship between different components of system. This diagram is very important to understand the overall concept of system. Architectural diagram is a diagram of system, in which the principal parts or functions are represented by blocks connected by lines that show the relationships of the blocks.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 4, April 2017

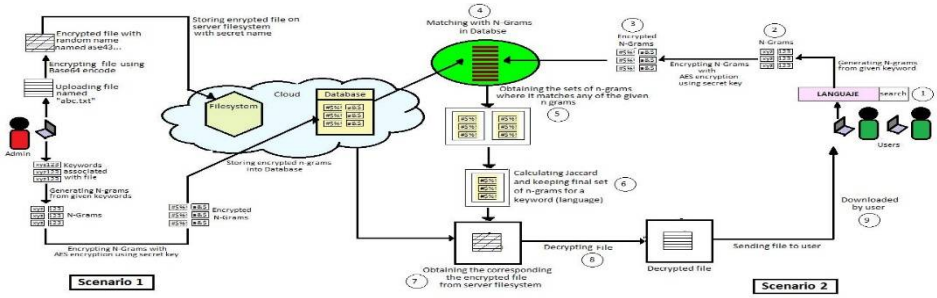



Fig. 2System Architecture

## V. IMPLEMENTATION

We designed the application based on system model. The system mainly focuses on user fast and easy retrieval of the file. The module program is a collection of features and interactive tool which are controlled by module operation and can be extended as various functions

This software is developed based on Code Igniter Framework. We apply JQuery as common library to develop Java Script along with HTML and CSS.

### 1. Searching

The searching is done by breaking the keywords typed into small parts containing 3 letters each called N-grams so that a spelling mistake can also lead to the result. And the N-Grams are stored in different table rather than a single table, thus, creating index for fast retrieval.

```
public function get_finfo($loc){
$query="select * from upl_files where f_loc='$loc';";
$records=$this->db->query($query);
returnUpl_files::instantiate($records);
}
 public function get_finfo_by_fid($fid){
 $query="select * from upl_files where f_id=$fid;";
 $records=$this->db->query($query);
 returnUpl_files::instantiate($records);
 }
public function get_finfo_by_uid($uid){ $query="select * from upl_files where f_u_id=$uid;";
 $records=$this->db->query($query);
returnUpl_files::instantiate($records);

}
```

### 2. Download

Display algorithm has been developed such that it first finds out the location and ex-tension of the file. That create a new file in temporary folder that copy the decrypted original file. This is done to protect the original file from any error.

```
public function download($fid){

$row=Upl_files::get_finfo_by_fid($fid);$original_filename="assets/files/".$row[0]->f_loc;

$ext=$row[0]->f_ext; $new_filename="assets/tmp/file_".rand(100,1000000).".".$ext;

$fp2=fopen($new_filename,'wb'); fputs($fp2,base64_decode(file_get_contents($original_filename)));

header('Content-Description: File Transfer'); header("Content-Type: application/".$ext);
```

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 4, April 2017

```
header('Content-Transfer-Encoding: binary');

header('Expires: 0');
header('Cache-Control: must-revalidate, post-check=0,

, pre-check=0'); header('Pragma: public');

header('Content-Disposition: attachment; filename="' .

$new_filenameheader("Content-Length: " . filesize($new_filename));

ob_clean();
flush();
readfile($new_filename);
fclose($fp2);

die;}
```

And the page session expires the files in the temporary folder are deleted automatically.

## 3. N-Gram

The 3 letter words created by breaking the keyword is called the N-gram. The N-Grams are stored in different tables thus creating index for fast access.

```
public function create_enc_ngrams($word,$n=2){
$ngrams=array();
$len=strlen($word);
for($i=0;$i<$len;$i++){
if($i> ($n - 2)){
$ng= '';
for($j=$n-1;$j>=0;$j--){
$ng.=$word[$i-$j];
}
$ngrams[]=$ng;
}
}
$enc_ngrams=array();
$i=0;
foreach($ngrams as $row){
$aes = new AES($row);
$enc_ngrams[$i++] = $aes->encrypt();
}
return $enc_ngrams;
}
```

## 4. JACCARD CALCULATION

The Jaccard index, also known as the Jaccard similarity coefficient is a statistic used for comparing the similarity and diversity of sample sets.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}.$$

(If $A$ and $B$ are both empty, we define $J(A,B) = 1$.)

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 4, April 2017

## VI. RESULT

We first present the results of each function of application based on our latest experiment. Our result Fig 3 shows the home screen.

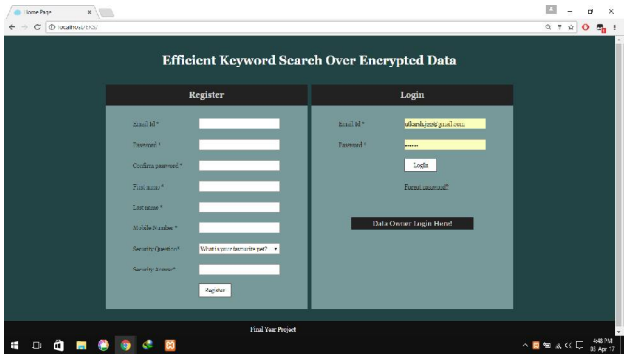



Fig. 3: Home screen loaded on page startup

Now there are 2 modules one for admin login and other for user login. In these modules the user/admin can upload any file, view all the files in case of admin and the user can view the files he has uploaded. Both the parties can delete any file displayed in the tabular view section and they can even search for a particular file through keyword.

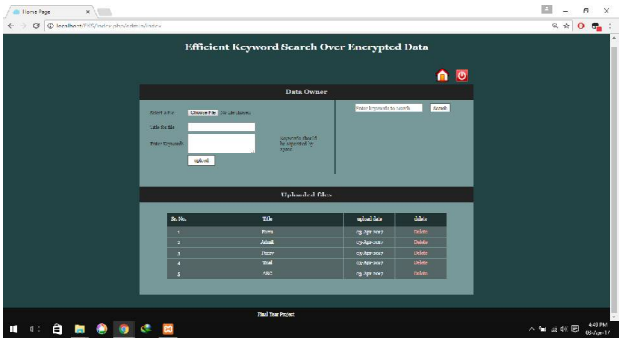



Fig 4: Admin/User Dashboard

Our result Fig 5 shows the Interface after the user/admin searches for a file. They input the keywords which are converted and encrypted into N-Grams and these are matched with previously stored N-Grams. The Interface provide the searched keyword along with the matched keyword and also displays the time required to execute the process.

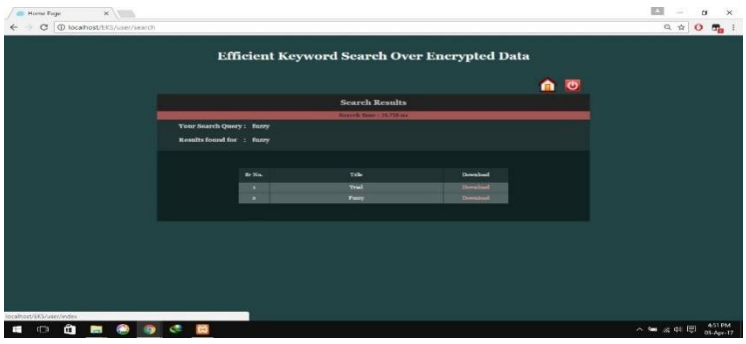



Fig 5: Search Result

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 4, April 2017

## VII. CONCLUSION

Our proposed model of using AES Encryption for security of the data is highly efficient. The search over the encrypted base-64 file has also been successfully executed. And the search optimization using Wildcard based technique along with Jaccard calculation for obtaining the similarity between the generated N-grams and previously encrypted N-Grams to obtain the desired files. And the N-grams are broken into small pieces of
3 letters each. Rather than storing all the N-Grams in single table we are storing the N-Grams in different table thus creating Indexes.

## REFERENCES

[1] J. Li Q. Wang and C. W. et al "Fuzzy keyword search over encrypted data in cloud computing" in 2010 Proceedings IEEE on INFOCOM . IEEE 2010 pp. 441-445.
[2] D. Boneh G. D. Crescenzo and R. O. et al. "Public key encyrption with keyword search" in Advances in Cryptology-EUROCRYPT 2004 ser. LNCS C. Cachin and J. Camenisch Eds. vol. 3027. Springer-Verlag 2004 pp. 506-522.
[3] C. Liu, L. Zhu, L. Li, and Y. Tan, "Fuzzy keyword search on encrypted cloud storage data with small Index," ICCIS 2011, 269-273, 2011.
[4] M. Chuah, and W. Hu, "Privacy-aware bedtree based solution for fuzzy multi-keyword search over encrypted data" ICDCSW 2011, 273-281, 2011.
[5] M. Chuah, and W. Hu, "Privacy-aware bedtree based solution for fuzW. Sun, B. Wang, N. Cao, M. Li, W. Lou, T. Hou, and H. Li, "Privacy-preserving multi-keyword text search in The cloud supporting similarity-based ranking," In ASIACCS 2013, 2013.zy multi-keyword search over encrypted data" ICDCSW 2011, 273-281, 2011.
[6] D. Song, D. Wagner, and A. Perrig, "Practical Techniques for Searches on Encrypted Data," in Proc. of IEEE Symposium on Security and Privacy'00, Pages 44-55, 2000.
[7] E.-J. Goh, "Secure Indexes," Cryptology ePrint Archive, Report 2003/216, 2003, http://eprint.iacr.org/.
[8] R. Brinkman. "Different Search Strategies on Encrypted Data Compared. Data-Centric Systems and Applications," Security, Privacy, and Trust in Modern Data Management, Part 3, Pages 183-196, 2007.
[9] https://en.wikipedia.org/wiki/Jaccard_index.
[10] S. Ji, G. Li, C. Li, J. Feng,"Efficient interactive fuzzy keyword search", in Proc. of WWW'09,2009.