IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper Volume-4, Issue-02 E-ISSN: 2347-2693

Fuzzy Keyword Search over Encrypted Stego in Cloud

TanmayDahake[1*], MirsohailShaikh[2], Vipul Khandeparkar[3],Neerav Panchal[4]and Gajanan Bherde[5]

[1*, 2,3,4,5]*Department of Computer Engineering,*
K.J. Somaiya College of Engineering, Mumbai University,India

www.ijcseonline.org

Received: Jan /19/2016 Revised: Feb/02/2016 Accepted: Feb/12/2016 Accepted: Feb/29/2016

Abstract**:** Today cloud is a relatively young and booming technological advancement. Cloud services are getting more and more popular these days, from online software to cloud storages like Google drive, one drive, Dropbox, etc. Smartphones and increasing internet connectivity are the main catalyst to this booming industry. The benefits that can be obtained from this technology are many like low cost, low maintenance and many more. Consumers now don't have to worry about the storage space, now they can subscribe to cloud storage and avail much more data storage capacity at low cost compared to actual hardware. But there are many problems too, security and privacy is one of the major one. In this paper we will propose a solution for this problem in the context of cloud storage. We propose usage of cryptography and steganography. Cryptography converts ordinary information (plain text) into something that doesn't make sense or incomprehensible information (cipher text). Steganography is concealment of information. Unlike cryptography it doesn't attract attention, it just appears to be a normal data but within it a sensitive data is hidden. The other problem is data retrieval or searching. We propose using fuzzy logic for this.Fuzzy keyword search incredibly improves framework ease of use by giving back the coordinating data when clients seeking inputs precisely coordinate the predefined keywords or the nearest conceivable coordinating data in view of keyword similitude semantics, when careful match comes up short.

Keywords: Cloud Computing, Fuzzy Keyword Search, Privacy, Cloud Security.

I. INTRODUCTION

As cloud computing becomes ubiquitous, lots of sensitive data is stored in the cloud, for e.g. electronic mails(E-mail), e-copy of passports and other important government documents, etc. Cloud relives the burden of storing and maintaining the uploaded data from the owners of the same. Hence, they can enjoy better quality data storage options whenever they require. Although, the fact that the data storage server(cloud server) and the data owner are not in the same domain may put the data at stake. It is recommended that sensitive data should be encrypted before uploading for data privacy and preventing unauthorized access. Although data encryption makes the utilization of the data a tedious task given that there would be huge amounts of uploaded data files. Furthermore, in cloud computing the users may share their data with huge number of other people. The individual user would be interested in certain specific data during a given session. One of the most widely used ways is retrieving the data using keyword based searching methodology instead of going through the list of all the data files in the cloud server. An example of search selective retrieval of data is Google search. Unfortunately data encryption leads to inability of using plain text keyword search in cloud computing. Excluding this encryption also requires the keywords to be kept exclusive as they may point out some sensitive information in the data file. The module of our project includes the security related to the actual confidential data on the cloud storage. Steganography and cryptography both are considered as methods to protect the digital information. Present work explored the application of cryptography as a wrapper of an audio file used to hide the actual data using method of steganography. Our approach encompasses AES and Base-64 encryption with audio steganography successfully increases the security and privacy while storing data in the cloud storage.

II. METHOD

We propose a two layered security for the data in cloud storage:

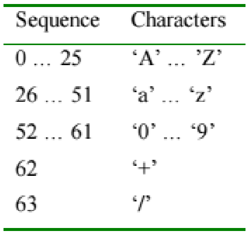
1. Encryption
2. Data hiding

In the proposed technique, first the audio file is sampled and then an appropriate bit of each sample is altered to embed the textual information. The audio file is then encrypted by base 64 technique to achieve better security in the audio steganography process.

1. Encryption:

- Our proposed system uses AES encryption in order to protect the set of key words generated using the gram based technique while storing the data on the cloud storage. The Advanced Encryption Standard (AES) is a symmetric-key block cipher published by the National Institute of Standards and Technology (NIST) in December 2001.AES is a non-Feistel cipher that encrypts and decrypts a data block of 128 bits. It uses 10, 12, or 14 rounds. The key size, which can be 128, 192, or 256 bits, depends on the number of rounds.[1]
- Base 64 encryption is used for the files. Base 64 data

representation is based on a 64-character alphabet. The alphabet is shown in Table.



Sequence	Characters
0 ... 25	‘A’ ... ’Z’
26 ... 51	‘a’ ... ‘z’
52 ... 61	‘0’ ... ‘9’
62	‘+’
63	‘/’

A binary file is a series of zero and ones. Group these zeros and ones in sets of 6, and then we get a number between 0 and 63 for each set. Converting them using the base64 alphabet allows converting binary zeros and ones into a compressed and human readable format. By compressed, we mean that data can represent six bits with one character. This is six times less than representing each bit with a ‘0’ or ‘1’ character. [2]

2. Data Hiding: After the encryption using the base-64 is done next comes the audio steganography. For audio steganography the technique used is the LSB technique which stands for least significant bit. Least significant bit (LSB) coding is the way to embed information in a digital audio file. By substituting the least significant bit of each sampling point with a binary message, LSB coding allows for a large amount of data to be encoded. As the data stored on cloud storage is bigger in size we found LSB to be more efficient.[3]

For data retrieval we propose using Gram based fuzzy keyword search. In this technique the gram of a string is a substring that can be used as a signature for efficient approximate search. While gram has been widely used for constructing inverted list for approximate string search, we use gram for the matching purpose. We propose to utilize the fact that any primitive edit operation will affect at most one specific character of the keyword, leaving all the remaining characters untouched. In other words, the relative order of the remaining characters after the primitive operations is always kept the same as it is before the operations. [4]

III. Workflow

Refer appendix 1 for workflow diagram.

1. User of our project has a registered account with us he can upload his document on same. The data user uploads goes through a couple of cryptographic procedure before it get stored in the database.
2. But before the cryptographic methods operating on the data, user is asked to enter some keyword related to data so that he could find it later through the feature of fuzzy keyword search.
3. Fuzzy keyword uses the technique of n-gram on the keywords and split it in subsequent substrings and stores in the database in encrypted form which is done using AES algorithm.
4. Form the encryption of the data file base-64 algorithm is used; after the data file is encrypted using the base-64 algorithm the next stage comes is audio steganography.
5. After the two step security, one of the encryption and second one of audio steganography the stego is stored in the database against the registered user.
6. After uploading of the data is done, later when user needs to download the data file he search for the file he uploaded earlier, searching method again uses the n-gram technique to search the filename entered.
7. First the encrypted file is extracted from the stego uploaded earlier and then decrypted to get the required data file for the user.

IV. Technologies Associated

1. HTML – Hypertext markup language, commonly referred as HTML is the most popular markup language used to develop web pages. In our project we have used HTML to create all the user interactions screens.
2. MySQL – MySQL is an open source relational database management system. We have used MySQL to manage the data related to user’s login information and all the data files he upload in his account.
3. PHP – It is a server side scripting language but is also used as a general purpose programming language. Different modules in our project like fuzzy search and cryptography related modules are implemented in it.
4. CSS – Cascading Style sheet is a style sheet language used to present the documents written in the markup language. In our project CSS is used for styling the user interface pages written in HTML.

V. Future Scope

1. Currently we have developed our project considering only personal computers and laptops, in future we are thinking of extending it to portable devices like smart phones and tablets.
2. Our project currently deals with the cloud storage security and ease of access, in future we can extend this for other storage platform like the database of a private organization.
3. Different type of files that can be stored using the proposed technologies can be extended to store multimedia files.

VI. CONCLUSION

In this paperinterestingly we propose to utilize both steganography and additionally cryptography at the same time. We were able to successfully implement these methods that we proposed using PHP. We utilized a propelled technique (i.e., gram-based techniques) to develop the capacity proficient fuzzy keyword set by exploiting two critical perceptions on the comparability metric of alter distance. After the fuzzy keyword set is created it is then secured through A.E.S encryption and put away for the further utilize. Client information is then secured through the two level of security of base-64 encryption and after that the audio steganography.

VII. REFERENCES

[1] Chih-Chung Lu and Shau-Yin Tseng , "Integrated Design of AES (Advanced Encryption Standard) Encrypter and Decrypter" ,IEEE International Conference on Application-Specific Systems, Architectures, and Processors (ASAP'02) 1063-6862/02

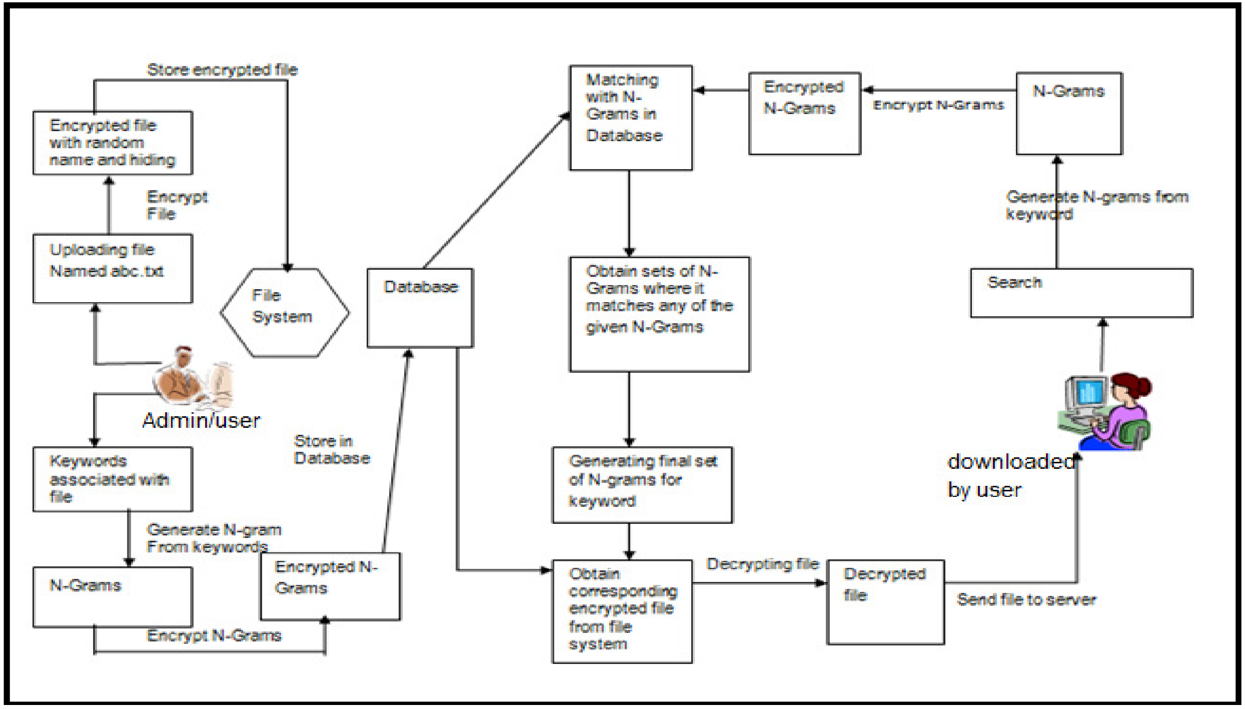
[2] Chih-Chung Lu and Shau-Yin Tseng , "Integrated Design of AES (Advanced Encryption Standard) Encrypter and Decrypter" ,IEEE International Conference on Application-Specific Systems, Architectures, and Processors (ASAP'02) 1063-6862/02

[3] "How to base 64" by Randy Charles Morin, www.kbcafe.com

[4] Gunjan Nehru, Puja Dhar, "A Detailed look of Audio Steganography Techniques LSB and Genetic Algorithm Approach", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 1, No 2, January 2012, ISSN (Online): 1694-0814

[5] Jin Li, Qian Wang, Cong Wang, Ning Cao, KuiRen†, and Wenjing Lou, "Fuzzy Keyword Search over Ecrypted Data in Cloud Computing," IEEE INFOCOM 2010,978-1-4244-5837-0/10

Appendix 1:


Store encrypted file
Encrypted file with random name and hiding
Encrypt File
Uploading file Named abc.txt
File System
Database
Admin/user
Keywords associated with file
Store in Database
Generate N-gram From keywords
N-Grams
Encrypted N-Grams
Encrypt N-Grams
Matching with N-Grams in Database
Encrypted N-Grams
Encrypt N-Grams
N-Grams
Generate N-grams from keyword
Search
Obtain sets of N-Grams where it matches any of the given N-Grams
Generating final set of N-grams for keyword
Obtain corresponding encrypted file from file system
Decrypting file
Decrypted file
Send file to server
downloaded by user