Fuzzy Keyword Search Over Encrypted Data using Cloud Computing

Teena Gupta, Rohit K.V.S.S

Abstract: ***"Cloud" is a collective term for a large number of developments and possibilities. It is not an invention, but more of a "practical innovation", combining several earlier inventions into something new and compelling. A cloud computing platform dynamically provisions, configures, reconfigures, and de provisions servers as needed [8]. Servers in the cloud can be physical machines or virtual machines. Security is a critical issue in cloud computing due to the variety of IT services that can be provided through a cloud environment. This paper focuses on the aspect of searching keywords over encrypted data while maintaining integrity of the data. Using a traditional algorithm like AES. Unlike traditional searching algorithms, here we try to implement a fuzzy logic which is based on a NLP technique called N gram. This fuzzy keyword searching significantly increases the efficiency and safety over cloud. This will keep the searching time efficient and acquire great results. The n-gram logic will be used to make sets of keywords which will used in the search implementation. To achieve more accurate results, Jaccard Coefficient will be used to find the similarity between the sets of keywords and rank them based on that. The purpose of this paper is to improve the traditional keyword search over encrypted data using cloud computing using advanced algorithms without compromising over security. Through rigorous security analysis, we show that our proposed solution is secure and maintains the privacy of the file server while efficiently using the fuzzy logic.***

Keywords: AES, N-Gram, Cloud Computing, Fuzzy Logic, Jaccard Coefficient

I. INTRODUCTION

In simple terms,cloud computing is a platform which delivers computing services that include storage, networks, software, analytics, developer tools and servers over the internet.They are essentially data centers which provide resources to users on-demand basis[1]. Cloud computing emerged in the wake of Web 2.0 ,it is divided in these three broad categories- IaaS(Infrastructure as a service), PaaS(Platform as a service) and SaaS(Software as a service). People these days rely heavily on cloud computing as it has become a concrete part of their life[2].

Cloud computing is a platform which delivers computing services that include storage, networks, software, analytics, developer tools and servers over the internet[3]. Because of its high efficiency and ease of access, large amount of sensitive data such as classified documents,healthcare records,financial documents etc are now being stored into cloud[4]. By doing this the data owner becomes vary of the security of the files which is being uploaded, which also raises the question of how the data can be outsourced effectively. An efficient file retrieval system through keyword based search instead of retrieving all the files associated[5]. This is similar to search engines like google, bing and yahoo etc.Here the files will be presented in encrypted format to provide maximum privacy and security. Now,the most primitive method is to get an exact string/keyword match.But in this paper we will be implementing fuzzy logic, so if a user encounters a typo or has insufficient knowledge about the file , it will still return the most closest result. Fuzzy search basically means to search for keywords whether they exactly match the user input keyword or not, it will provide you with the nearest keyword that matches your input[6]. It is the base for many well known search engines out there, for example : Google, Bing etc. In this paper we try to implement this over encrypted data using cloud computing. This can be achieved using various methods like Levenshtein distance also called edit distance or wild-card based searching[7]. We will be using the NLP technique-N grams.This will be used for the separation of input keyword into n grams or sets. Hence, in this paper we try to solve the problem of effective yet privacy preserving keyword search using nlp techniques and proper encryption schemes[1,8].

II. LITERATURE SURVEY

A. Fuzzy Keyword Search over Encrypted Data in Cloud Computing

In this paper , the author has come up with a unique technique (wild card based technique) to construct the fuzzy keyword sets that work on the notable observation of the similarity metric of edit distance .this method is used to support an efficient and privacy protecting fuzzy keyword search for encrypted data stored in cloud.This paper sets a good understanding of what fuzzy search is[1,9].

B. Implementation of Fuzzy keyword search over encrypted data in cloud computing

The main objectives of this study is-

- to make a privacy preserving fuzzy search for achieving effective usage of remotely stored encrypted data[2,10].

Manuscript received on 24 September 2022 | Revised Manuscript received on 28 September 2022 | Manuscript Accepted on 15 October 2022 | Manuscript published on 30 October 2022.

* Correspondence Author (s)

Teena Gupta*, UG Scholar, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur (Tamil Nadu), India. E-mail" tu2316@srmist.edu.in

Rohit K.V.S.S, UG Scholar, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur (Tamil Nadu), India. E-mail: rk1627@srmist.edu.in

- Another objective is to design an advanced search mechanism for constructing storage efficient fuzzy-keyword sets based on similarity metric edit distance

C. Secure Cloud Storage and quick keyword based retrieval system[3,11]

The main objectives of this study is-

- Various encryption techniques for storing data in cloud
- Keyword based file retrieval system
- Keyword extraction

III. PROPOSED WORK

The purpose of the project is to create a system through which file retrieval of encrypted data can be done using Fuzzy keywords while still keeping the file data secure[4,12]. The project mainly has three modules Admin, User and Database. The Admin here uses AES encryption technique to encrypt the files before uploading them on the cloud, admin can add, update and delete the files. Then we are using N-gram technique to store the file keywords in the database[13]. The user here has to give the personal details to the admin and register first so to have access of the encrypted data, after registering with the admin the user can enter a keyword here the fuzzy keyword search would help the user to get the most relevant files related to the search keyword. As the whole data is encrypted and stored on the cloud the data is secure and the confidentiality is maintained[14].

A. N-grams

N grams[15,16] is just another way to efficiently construct a fuzzy set.Here the keyword is divided to grams that can be used as a signature for efficient approximate search().Although n grams have been widely used for building inverted list for approximate word search, we use it for keyword matching purpose[1].We propose to make use of the fact that any edit operation will affect one specific character of the keyword, leaving all the remaining characters unaffected. By doing this, the relative order of the characters left after the operations is always kept same as it were before performing them. For example, the gram-based fuzzy set SCASTLE, 1 for keyword CASTLE can be constructed as {CASTLE, CSTLE, CATLE, CASLE, CASTE, CASTL, ASTLE}.

B. Encryption techniques

AES algorithm (also recognized as Rijndael algorithm) is a symmetrical block cipher algorithm which takes the plain text in block sizes of 128 bits and converts them into cipher text using varying key sizes of 128,192 and 256 bits[7].It uses a substitution-permutation method, along with multiple rounds to construct cipher text. Based on the key size, number of rounds will be decided. Every round requires a round key .The user inputs one key which is expanded to obtain keys for each round.

C. Database

The user details are stored in the database. The results and the admin related details are stored in the database.

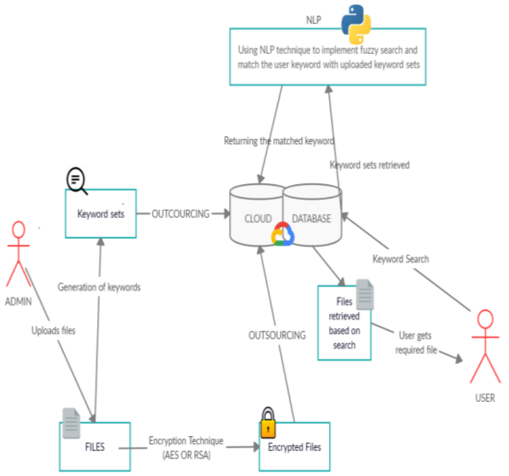

NLP
Using NLP technique to implement fuzzy search and match the user keyword with uploaded keyword sets
Returning the matched keyword
Keyword sets retrieved
Keyword sets
OUTCOURCING
CLOUD
DATABASE
ADMIN
Generation of keywords
Uploads files
Keyword Search
Files retrieved based on search
User gets required file
USER
OUTSOURCING
FILES
Encryption Technique (AES OR RSA)
Encrypted Files

Fig. 1. Architecture Diagram

IV. UML DIAGRAMS

Unified Modeling Language or UML diagrams is a modeling approach through which we can visually represent the working flow of the software. It is a business process modeling technique which is used to design a system. UML diagrams are divided into two basic categories Behavioral and Structural UML diagrams. This distinction is done on this bases of what the UML diagram depicts about the system. One describes the behavior of the system and different components of it while other depicts the structure and how all the components will be inked to each other. Some basic UML diagrams used for visual representation of the software are Use Case diagram, Activity diagram, Sequence diagram, Class diagram are many more.

A. Use-Case Diagram

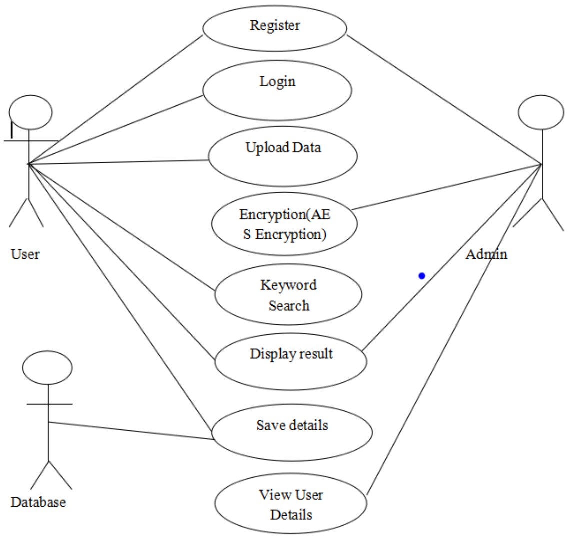


Fig. 2. Use case diagram for the system

The above figure represents use case diagram of the proposed system in which user, admin and database roles are shown. It shows the various functions that every role can perform in the system.

Retrieval Number: 100.1/ijeat.A38421012122
DOI: 10.35940/ijeat.A3842.1012122
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

B. Sequence Diagram

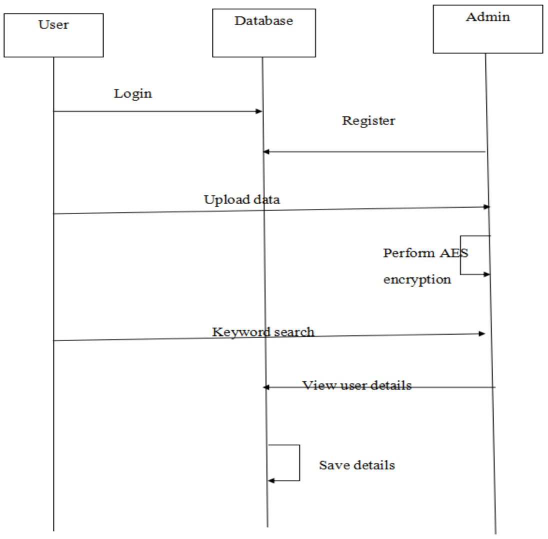


Fig. 3. Sequence Diagram

IT shows the interaction between a set of objects, through the messages that may be dispatched between them. This diagram shows the course of events occurring in the application and the interactions between different objects and actors.It also shows the messages in-between them, here we start with the registration on the admin side ,so that we can login later from user side.After uploading the data,AES is then performed on the data.Now the use can perform keyword search and retrieve the file.

C. Activity Diagram

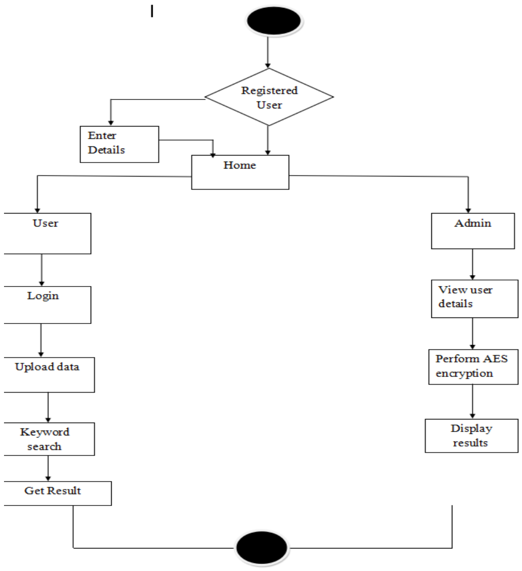


Fig. 4. Activity Diagram

Activity Diagram helps in understanding the work flow of the program through a defined start point to an end point, it’s a behavior diagram which is similar to state diagram. The key feature of the activity diagram is that it can show conditional activities.The above Activity Diagram describes the work flow of our system, which starts by the registration of the user, entering the asked details then directing the user to the home page. On the Home page, there are two ways to proceed either User or Admin. If User then login with credentials registered then download the data, and if Admin then can view the User Details, and upload the files by encrypting them by AES Encryption. After that both User and Admin can end the program

D. Data Flow Diagram

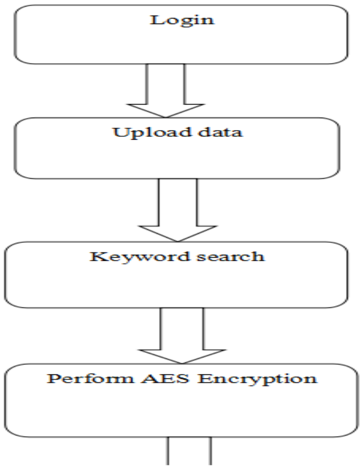


Fig. 5. Data flow diagram

V. MODULE DESCRIPTION

This implementation project consists of the following 4 modules:

A. User /Admin Registration

To use the fuzzy keyword search over encrypted files, the Admin and User both needs to register on the site. For registration the details needs to be filled in and the Admin needs to generate a 16-digit secure key, then the file can be uploaded to the cloud which will get encrypted using AES technique. Further to retrieve the desired file the user also needs to register and has to get the 16-didgit secure key generated by the admin. As the user enters the secure key, the file then will be automatically get decrypted using AES technique and downloaded for the user.

B. File Encryption

Here the files which were uploaded by the user get encrypted and uploaded in the cloud ,where they are stored in 4 different file formats .The keywords are stored in encrypted format as well.

C. Fuzzy Logic

Now for the fuzzy logic, we will be using the n gram technique. Here n grams are generated from the given keywords and then encrypted . These encrypted n grams are stored in the cloud. When the user wants to retrieve a file , he will be asked to enter the associated keyword , which will again be converted to n grams and encrypted , these will be matched and the relevant file will be displayed.

D. File Retrieval/Decryption

Now to retrieve the file, the user needs to enter the secret 16 digit key which he entered during the registration, using that the files are decrypted and downloaded on to the local file server.

VI. RESULTS DISCUSSION

The user uploads files which are encrypted and stored in cloud.

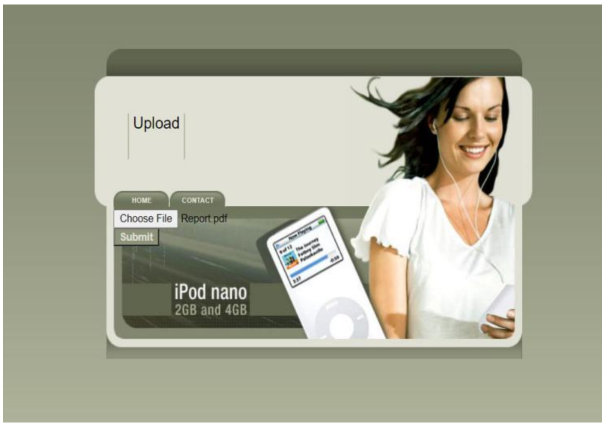
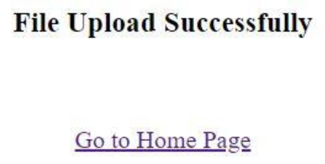


Fig. 6. File Upload page



File Upload Successfully

Go to Home Page

Fig. 7. File upload confirmation

The files are stored in four different file format In the database.

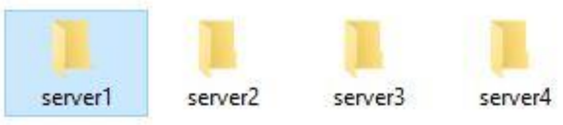


Fig. 8. File stored format

Now the files can be retrieved after the secret 16 bit key has been entered.

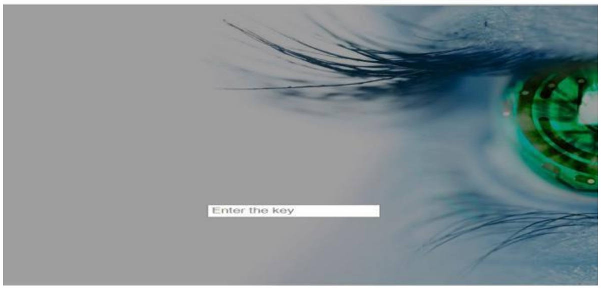


Fig. 9. File Download page

VII. CONCLUSION

In this paper we try to achieve the goal of an efficient file retrieval system using fuzzy logic and maintaining the security of the files at the same time by encrypting them with a secure encryption algorithm.

REFERENCES

1. J. Li, Q. Wang, C. Wang, N. Cao, K. Ren and W. Lou, 2010, "Fuzzy Keyword Search over Encrypted Data in Cloud Computing," , Proceedings IEEE INFOCOM, San Diego, CA, USA, pp. 1-5, doi: 10.1109/INFCOM.2010.5462196. [CrossRef]
2. Shekokar, N., Sampat, K., Chandawalla, C. and Shah, J.,Shekokar, N. et al., 2015, "Implementation of Fuzzy Keyword Search over Encrypted Data in Cloud Computing", Procedia Computer Science, 45, pp. 499-505. doi: 10.1016/j.procs.2015.03.089. [CrossRef]
3. Songfeng Lu; Abdulruhman I Ahmed AbomakhelbSecure, 2017, Cloud Storage and Quick Keyword Based Retrieval System , 2017 International Conference on Computing Intelligence and Information System (CIIS): https://ieeexplore.ieee.org/document/8327736 .
4. Yadav, Manish & Gugal, Drishti & Matkar, Shivani & Waghmare, Sanket., 2019, Encrypted Keyword Search in Cloud Computing using Fuzzy Logic, 2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT) 1-4. 10.1109/ICIICT1.2019.8741364. [CrossRef]
5. S.Amudha, M.Murali, 2020,'Deep Learing based energy efficient novel scheduling algorithms for body-fog-cloud in smart hospital', *Journal of Ambient Intelligent Humanized Computing*, https://doi.org/10.1007/s12652-020-02421-0 [CrossRef]
6. Muhammad Aliyu, M.Murali, Abdul Salam Y.Gital, Souley Boukari, 2020,' Efficient Metahueristic Population Based and Deterministic Algorithm for Resource Provisioning using Ant Colony Optimization and Spanning Tree', *International Journal of Cloud Applications and Computing*, 10(2), 1-21. [CrossRef]
7. Muhammad Aliyu, M.Murali, Zuopeng Justin Zhang, Abdul Salam Y.Gital, Souley Boukari, Yongbin Huang, Ismail Zahraddeen Yakubu, 2021, 'Management of cloud resources and social change in a multi-tier environment: A novel finite automata using ant colony optimization with spanning tree', *Technological Forecasting & Social Change*, 166(2021) 120591 [CrossRef]
8. M.Murali, 2015,' Principal Component Analysis based Feature Vector Extraction', *Indian Journal of Science and Technology*, 8(35), 1-4 [CrossRef]
9. Ankita Sadh, M.Murali, 2016, 'Wireless Sensor Data Access Through Mobile Cloud Computing', *International journal of Control Theory and Applications*, 9(15), 7325-7331
10. Ankita Sadh, M.Murali, 2016, 'Wireless Sensor Data Access Through Mobile Cloud Computing', *International journal of Control Theory and Applications*, 9(15), 7325-7331
11. Shaik Saleem. M, Murali, 2018, 'Privacy preserving public auditing for data integrity in cloud', *Journal of Physics: Conf. Series 1000*, doi: 10. 1088/ 1742-6596/ 1000/ 1/012164 [CrossRef]
12. M.Murali, R.Srinivasan, 2015,'Cached Data Access in MANET employing AODV protocol', IC4-2015(IEEE international conference), Indore, Madhya Pradesh. [CrossRef]
13. A.Venisha, M.Murali, 2019, 'Discovering the Trustworthy Cloud Service provider in Collaborative Cloud Environment', *International Journal of Engineering and Advanced Technology*,9(252), 360-367
14. A.Venisha, M.Murali, 2019, 'A Conception for identifying trust service providers in collaboration cloud computing',*International Journal of Recent Technology and Engineering*, 8(254), 110-116 [CrossRef]
15. Rentachintala Kasyap, M.Murali, 2020, 'Privacy, Data Management and Access Control in Smart Meters: A Survey', *European Journal of Molecular & Clinical Medicine', 7(5), 1630-1645*
16. J.Shobana, M.Murali, 2021, Abstractive Review Summarization based on Improved Attention Mechanism with Pointer Generator Network Model, Webology, Volume 18, Number 1, DOI: 10.14704/WEB/V18I1/WEB18028 [CrossRef]

AUTHORS PROFILE

Teena Gupta, UG Scholar, Department of Computer CSE, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu. Currently working as a Technology Program Analyst II at Fiserv. With great interpersonal skills, and problem capabilities have worked on a lot of different projects and have also contributed to various research works. In current job role at a fintech company, got exposure to learn about the financial world. Working in a product company allowed me to gain insight into technologies like JAVA, SQL, and DevOps. Consistently strive to take on my new challenges and explore the new advancing technologies in the industry. Committed to bringing better and advanced solutions for business problems.

Rohit K.V.S.S, UG Scholar, Department of Computer CSE, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, Ambitious and career-oriented, finished BTech with distinction and took up a campus placement. Profound knowledge in cryptography and blockchain, and hands-on experience in the Ethereum platform. Experienced in web programming and full-stack development. Worked on multiple frameworks like Angular, React and Angular.js along with backend development in .NET. Recent indulgence in the field of finance and seeking to pursue a career in computational finance and business analytics. Apart from maintaining an excellent academic record, has also participated in various extracurricular activities. Expanded knowledge about new and upcoming trends in computing industries by working on various projects and researching about different domains like cloud computing, and image processing.