Gram-Based Fuzzy Keyword Search Over Encrypted Data in Cloud Computing

Vikas Gupta, Department of Computer Science & Engineering, RDEC, Ghaziabad. v.gupta@gmail.com

Abstract

The market for portable devices has recently seen tremendous growth, and with it, the use of cloud computing. The cloud's primary computers contain a great deal of sensitive information. File searching is made more difficult by the fact that these files are often encrypted before uploading in order to protect privacy. While earlier iterations of cloud computing's searchable encryption algorithms did make it possible to safely search encrypted data using keywords, these methods only worked for precise matches and would not work with variations in spelling or morphology. Fuzzy keyword search over encrypted cloud data is solved in this study. K-grams are used to generate imprecise outcomes. Two independent servers, unable of exchanging data with one another, are used for the purpose of security. The results of our experiments validate the efficacy and scalability of our technology in dealing with massive volumes of encrypted information.

Keywords: cloud computing,encryption algorithms,Fuzzy,K-grams

Introduction

Over the course of the last few years, a huge number of individuals have begun to use cloud computing services for their respective professions. People have the ability to save, access, and share their information at any time and from any location thanks to cloud computing. As a consequence of this, an increasing amount of sensitive information is being kept on the cloud. Users have the expectation that the cloud service provider would not only ensure their privacy and security but also ensure that their activities are completed efficiently in such an open environment.The term "cloud storage" refers to a type of online storage in which users upload their files, and the data they upload is then kept on various virtual servers, which are often hosted by third parties, rather than on dedicated servers.

Only authorised users, such as the owners of the data, are able to view the information that has been saved. In order to better safeguard their data, users often encrypt not only the contents of their files but also the names of their files before uploading them to the cloud. This makes it more difficult for the cloud storage provider to sift through the data.In the most recent years, searchable encryption methods have been developed as a solution to these issues [1-10].

These techniques, on the other hand, are not scalable since they are too sluggish to be applied to a big dataset for any reason. In addition to this, users often make typographical mistakes or use morphologically different forms of the same word. In light of this, the search service for cloud storage need to allow fuzzy searching.The purpose of this research is to offer a fuzzy keyword ranked search engine that is based on k-grams and operates on encrypted cloud data in order to overcome these challenges.

Two distinct servers—a search server and a storage server—are used by our organisation in order to enhance the safety of our system. Once the search server has been hacked, the attacker will be unable to utilise the file access pattern to determine the related document that is stored in the storage server. This is because the attacker will not be able to access the file. When we generate fuzzy keyword sets, we make use of k-grams, and when we measure the similarity between keywords, we utilise the Jaccard coefficient. The elimination of keywords having a Jaccard coefficient that is lower than our threshold value is done in order to prevent the process of enumerating all fuzzy keywords and, as a result, to reduce the search space. The determination of this threshold value is accomplished via the many experiments that are detailed in Section 5. Our suggested weighted ranking mechanism is used to determine the order in which search results are displayed. If a user uses our system, they will be able to do a fuzzy keyword search in an encrypted environment in a safe and efficient manner.

A platform that provides computing services such as storage, networks, software, analytics, developer tools, and servers via the internet is what we mean when we talk about cloud computing..In their most basic form, they are data centres that provide resources to

consumers on an as-needed basis [1]. The concept of cloud computing arose in the aftermath of Web 2.0, and it may be broken down into three basic categories: infrastructure as a service (IaaS), platform as a service (PaaS), and software as a service (SaaS).
People in this day and age are very reliant on cloud computing since it has become an integral component of their everyday lives [2].
The manuscript was received on September 24th, 2022; the revised manuscript was received on September 28th, 2022; the manuscript was accepted on October 15th, 2022; and the manuscript was published on October 30th, 2022.
* Exchange of letters Teena Gupta, Junior Undergraduate Scholar in the Department of Computer Science and Engineering at SRM Institute of Science and Technology in Kattankulathur, Tamil Nadu, India, is the author or authors of this scholarly work. A message sent to tu2316@srmist.edu.in
Rohit K.V.S.S is a student at the undergraduate level attending the SRM Institute of Science and Technology in Kattankulathur, Tamil Nadu, India. He is a member of the Department of Computer Science and Engineering. Send an email to rk1627@srmist.edu.in.
Cloud computing is a platform that provides computing services such as storage, networks, software, analytics, developer tools, and servers over the internet [3]. Cloud computing is also known as internet-based computing. As a result of the cloud's great efficiency and the simplicity with which it can be accessed, a significant quantity of sensitive data, including classified papers, healthcare records, financial information, and other similar documents, is now being kept there [4]. When this is done, the owner of the data is uncertain about the safety of the files that are being uploaded, which also begs the issue of how the data can be outsourced in an efficient manner. A system that retrieves files efficiently by using a search that is based on keywords rather than obtaining all of the files linked with the search [5]. This operates in a manner comparable to search engines such as Google, Bing, Yahoo, and others.In order to provide the highest possible level of confidentiality and safety, the files will be delivered in an encrypted manner here.
In this study, however, we will be implementing fuzzy logic, which means that even if a user meets a mistake or has inadequate understanding about the file, it will still produce the answer that is the closest to the original description. In its most basic form, fuzzy search does a search for keywords, regardless of whether or not those keywords precisely match the user's input keyword. It then provides you with the term that is the closest match to the one you entered [6]. Many of the most well-known search engines in existence, such as Google, Bing, and others, begin with this as their foundation. Our goal in this work is to demonstrate how cloud computing may be used to do this over encrypted data. Numerous techniques, such as the Levenshtein distance, which is sometimes referred to as the edit distance, or wild-card based searching, may be used to accomplish this goal [7]. We are going to make use of the NLP method known as N grammes.When the input keyword is separated into n grammes or sets, this will be utilised for the separation process. By using nlp methods and appropriate encryption schemes, we endeavour to provide a solution to the challenge of conducting an efficient keyword search that does not compromise the user's privacy in this study [1,8].

Proposed Work

The project's objective is to develop a system that will allow for the recovery of encrypted data files via the use of fuzzy keywords, while at the same time ensuring that the data contained inside the files remains safe [4,12].A primary focus of the project is comprised of three modules: Admin, User, and Database. Before uploading the data to the cloud, the administrator utilises the Advanced Encryption Standard (AES) encryption technology to encrypt them. The administrator has the ability to add, edit, and remove the files. Our next step is to make use of the N-gram approach in order to save the file keywords in the database

[13]. The user is required to provide the administrator with their personal information and register in order to get access to the encrypted data. Once the user has registered with the administrator, they are able to input a keyword, and the fuzzy keyword search will assist them in obtaining the files that are most pertinent to the search phrase.
As a result of the fact that the whole data set is encrypted and kept in the cloud, the data is protected and the secrecy is preserved.

N-grams

The use of N grams is only an additional method for effectively constructing a fuzzy set.In this case, the term is broken down into grammes, which may be used as a signature for an effective approximation search().While n grammes have been used extensively for the purpose of constructing an inverted list for the purpose of approximation word search, we utilise it for the purpose of keyword matching [1].Taking use of the fact that every edit operation will only alter one particular character of the keyword, while leaving all of the other characters intact, is the strategy that we propose to implement. Through this method, the relative order of the characters that are left after the operations is always maintained in the same manner as it was before the operations were carried out.

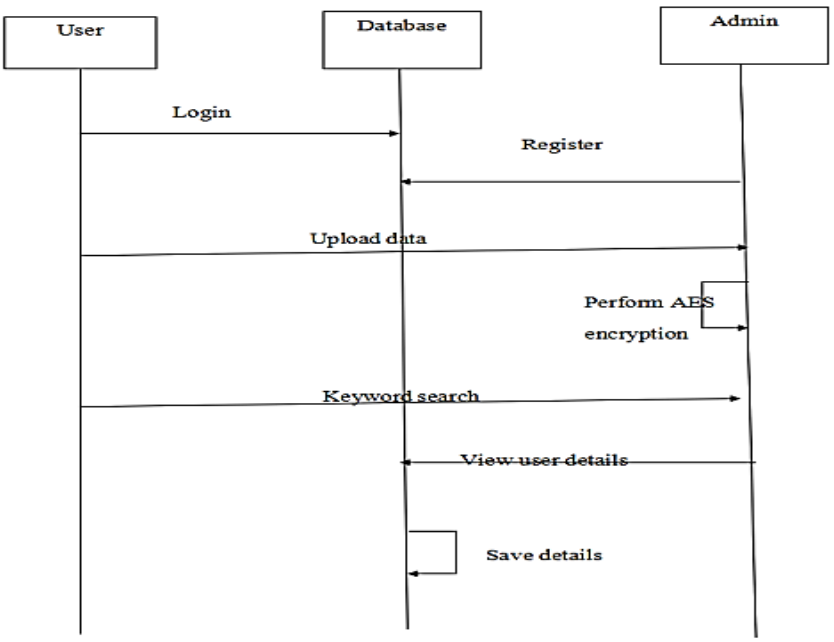

Fig. 2. Sequence Diagram

Conclusion

We make an effort in this article to accomplish the objective of developing a file retrieval system that is both effective and efficient via the utilisation of fuzzy logic, while simultaneously ensuring the safety of the information through the utilisation of a secure encryption technique**.**

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,'DeepLearing 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, SouleyBoukari, 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, SouleyBoukari, Yongbin Huang, Ismail ZahraddeenYakubu, 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. AnkitaSadh, M.Murali, 2016, 'Wireless Sensor Data Access Through Mobile Cloud Computing', International journal of Control Theory and Applications, 9(15), 7325-7331
10. AnkitaSadh, M.Murali, 2016, 'Wireless Sensor Data Access Through Mobile Cloud Computing', International journal of Control Theory and Applications, 9(15), 7325-7331
11. ShaikSaleem. 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. RentachintalaKasyap, 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]