Fuzzy Searching System over Encrypted library Data in Cloud

## Abstract

In this project, we proposed a secure cloud based library search engine for colleges. Every college will be able to upload their e-books, journals, notes, assignments etc on cloud server in encrypted format. At the same time the system will extract keywords from uploaded document using RAKE algorithm and store that keywords on cloud server in encrypted format using homomorphic encryption technique. To improve security of the system, we proposed modified AES algorithm for document encryption and Caesar algorithm for keyword encryption. Along with security of the documents and keywords, we proposed a fuzzy searching system for students. Using fuzzy searching system, the students will be able to search relevant books from uploaded books and other documents on cloud.

## Introduction

Cloud storage has nowadays become one of the most popular data storage solutions, where users can outsource their data to the cloud server for the low cost and convenient access. According to the report from Research and Markets , cloud storage market in 2017 is US$25.171 billion, and it will reach US$92.488 billion by 2022 at a compound annual growth rate of 29.73% from 2017 to 2022. However, the privacy and security issues are still the main challenges concerned in cloud storage. The cloud storage data breaches happened from time to time due to attacks, malfunctions or misconﬁgurations, such as Apple iCloud celebrity leak , Dropbox password leak and medical data leak on Amazon . In this case, it would be wise to encrypt the data before uploading them to the cloud server. Meanwhile, in order to efficiently retrieve the data on demand, an encrypted index should be associated with the underlying encryption scheme, which results in the so-called searchable encryption. In our proposed system, we proposed homomorphic encryption algorithm to encrypt keywords on cloud. In our project we focus on security of documents as well as keywords of the documents. Along with storage we proposed fuzzy based search engine to improve searching availability for students.

## Literature Review

In most SE-based cloud storage systems [1], [2], the cloud server is assumed to be honest but-curious. That is, the cloud server would exactly follow the speciﬁc protocol. Under this trust assumption, everything goes well. For example, the user issues (encrypted) search query to the cloud server, and the latter will return the valid and complete search results to the user. However, this trust assumption is not always true in reality. In other words, the cloud server would return an invalid or incomplete search result to the user, due to the missing of the data [3], [4] or monetary reasons. In this regard, the veriﬁability of the search result in the cloud storage becomes a requirement.

Chai and Gong [5] ﬁrstly investigated the veriﬁability problem in searchable encryption. Since then, many research efforts have been dedicated to design veriﬁable SE in symmetric key setting [6]–[7] or public key setting [8]–[9]. However, almost all existing veriﬁable SE schemes suffer from at least one of the following disadvantages: i) some schemes lack the functionality simultaneously supporting query operators “OR”, “AND”, “∗” and “?”, which are considered as the basic search ability of the traditional database system; ii) other schemes supporting the above four kinds of query operators usually require some time-consuming operations, such as bilinear maps, which is against the trend that the resource-constrained devices, such as smart phones, are becoming the main devices for people to access internet [10].

## Problem Statement

In existing system, AES algorithm is used for document encryption. AES algorithm is a unbreakable encryption algorithm but if any user leak document key, there is a possibility of document leakage. As attacker will be able to apply AES decryption code easily (this code is available anywhere on internet). Therefore to increase security of the system in case of key leakage we proposed modified AES algorithm.

## Proposed Methodology

## Modules

* Admin panel
  + Login
  + Register staff
  + Register library admin
  + View uploaded study material
  + View staff
  + View students
  + Approve pending students
* Staff panel/ Librarian panels
  + Login
  + Upload book/journals/ notes/ videos etc
  + Search any document
  + Decrypt document using secrete key
* Student panel
  + Login
  + Search documents by using keywords
  + Get relevant documents
  + Get otp on email
  + Specify otp
  + If otp verified, get decrypted document
* Keywords extraction and encryption
  + Upload document
  + Read complete document into string
  + Apply RAKE algorithm
  + Extract keywords
  + Generate secrete key k
  + Encrypt keywords using Caesar algorithm using k
  + Store encrypted keywords on cloud server
* Document encryption
  + Upload document
  + Generate secrete key k1, k2, k3, k4
  + Split document into four parts
  + Reverse the bytes of parts
  + Encrypt documents using AES and keys k1,k2,k3, k4 respectively
  + Combine all parts to get encrypted document
  + Store encrypted document on cloud server
* Fuzzy Searching
  + In computer science, **fuzzy** string **matching** is the technique of finding strings that match a pattern approximately (rather than exactly). In another word, **fuzzy** string **matching** is a type of **search** that will find matches even when users misspell words or enter only partial words for the **search.**
* **References**

1. C. Zuo, S. Sun, J. K. Liu, J. Shao, and J. Pieprzyk, “Dynamic searchable symmetric encryption schemes supporting range queries with forward (and backward) security,” in ESORICS, 2018, pp. 228–246.
2. H. Li, D. Liu, Y. Dai, T. H. Luan, and S. Yu, “Personalized search over encrypted data with efﬁcient and secure updates in mobile clouds,” IEEE Trans. Emerging Topics Comput., vol. 6, no. 1, pp. 97–109, 2018.
3. H. Blodget, “Amazon’s cloud crash disaster permanently destroyed many customers’ data,” Online, Apr. 2011, http://www.businessinsider.com/ amazon-lost-data-2011-4.
4. R. Jennings, “Oops: Google ‘loses’ your cloud data (sky falling; ﬁlm at 11),” Computerworld, Aug. 2015, https://www.computerworld.com/article/2973600/cloud-computing/ google-cloud-loses-data-belgium-itbwcw.html.
5. [5] Q. Chai and G. Gong, “Veriﬁable symmetric searchable encryption for semi-honest-but-curious cloud servers,” in IEEE ICC, 2012, pp. 917– 922.
6. [6] J. Wang, X. Chen, H. Ma, Q. Tang, J. Li, and H. Zhu, “A veriﬁable fuzzy keyword search scheme over encrypted data,” Journal of Internet Services and Information Security, vol. 2, no. 1/2, pp. 49–58, 2012.
7. W. Ogata and K. Kurosawa, “Efﬁcient no-dictionary veriﬁable searchable symmetric encryption,” in Financial Cryptography and Data Security, 2017, pp. 498–516.
8. [8] W. Sun, B. Wang, N. Cao, M. Li, W. Lou, Y. T. Hou, and H. Li, “Veriﬁable privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking,” IEEE Transactions on Parallel and Distributed Systems, vol. 25, no. 11, pp. 3025–3035, 2014.
9. S. Jiang, X. Zhu, L. Guo, and J. Liu, “Publicly veriﬁable boolean query over outsourced encrypted data,” IEEE Transactions on Cloud Computing, vol. PP, no. 99, pp. 1–1, 2017.
10. [23] J. Steven, “Internet stats & facts for 2019,” Online, Dec. 2018, https: //hostingfacts.com/internet-facts-stats/.