Secure Keyword Search Using Dual Encryption in Cloud: An Approach

Husna Tariq
Department of Computer Science, Jamia Hamdard,
Hamdard Nagar, New Delhi, 110062, India

Dr. Parul Agarwal*
Department of Computer Science, Jamia Hamdard,
Hamdard Nagar, New Delhi, 110062, India

*Corresponding Author

Abstract

This paper deals with the secure searching, storage and retrieval of user data in the cloud system. Various services of cloud, security issues and security requirements of cloud data are discussed. We have used fuzzy keyword searching scheme to search and retrieve the encrypted file by employing wildcard technique. We present a new approach of dual encryption system based on authentication of the server to provide stronger security to the existing fuzzy keyword searching schemes. We have integrated symmetric and asymmetric encryption algorithms to enhance data security. This work mainly focuses on authentication of the server so as to improve the security system and protect sensitive user’s data from unauthorized disclosure.

Keywords: Dual Encryption; authentication; RSA; AES; fuzzy keyword search.

1. INTRODUCTION

The cloud computing provides the facility to the variety of applications operating over thousands of computers and servers to concurrently access the services through internet. With the evolvement of cloud computing now it has become easier for users
to store, retrieve and share their data among themselves. It offers various benefits to users as well as to service providers. It provides flexibility to work from anywhere at any time. It also provides low-budget services, updates software automatically, raises collaboration amongst users and service vendors and much more.

The most extensively adopted application of cloud computing is cloud storage. A tremendous amount of information is being stored by users on cloud servers every day. This information needs protection from different kinds of cyber threats.

To maintain data confidentiality and for secure storage, various types of encryption algorithm are used for protecting information from unauthorized disclosure. However, searching over encrypted data was difficult to attain. Therefore, keyword based searching has been introduced where the desired file is retrieved after searching for it by providing the corresponding keyword.

The traditional methods of keyword searching were limited to the exact keyword search. But nowadays, many researchers have implemented fuzzy keyword searching in which the encrypted file is retrieved when the keyword matches exactly or when it is slightly misspelled and preserving the privacy of keywords at the same time. Shekokar et al. [18] have implemented fuzzy keyword searching by using wildcard technique in a semi-trusted server.

Authentication and access control mechanisms perform crucial role in supporting security and data protection in cloud system [13]. Access control mechanisms are practiced in database management system to provide protection to cloud data. Access control mechanisms work accurately only when the server is trusted [25]. But when we have an untrusted server then access control mechanisms fails to provide the required security [25]. Therefore, the cloud server should be trustworthy and authenticated so that users can securely store and retrieve their data.

So the goal of this paper is to focus towards the implementation of authentication at server’s side. Users must be able to verify whether his downloaded data comes from an authentic source or not. So we have suggested an idea of dual encryption that combines the advantage of symmetric and asymmetric encryption algorithm to provide additional security to the fuzzy keyword searching technique in the cloud system. The proposed work adds RSA algorithm to encrypt the message with the private key of the server before the user downloads his file so that he can verify whether the message comes from an alleged source or not. First of all, this paper addresses various cloud services, Issues related to cloud security and safety requirements in section 1. Section 3 introduces with the literature survey and the works done in this field till now. We have also presented a brief description of cryptographic algorithms and technique used for implementing fuzzy keyword searching in section 4. Section 5 defines the problem statement, and section 6 and section 7 presents proposed work and implementation details providing authentication service to the existing scheme. Section 8 provides the conclusion to the paper.
2. CLOUD SERVICES, ISSUES AND REQUIREMENTS

A security breach could happen at any of the entity involved in the cloud network. According to [9], the cloud comprises of three entities:

- **User**: The one who is interested in utilizing the services of cloud service providers. It can be individual customer, firm or any organization. They depend on the cloud to store their data.
- **Cloud Service Provider**: This entity has authority to manage the servers dispersed at different locations, and it controls whole cloud computing system since it provides crucial resources and expertise to users.
- **Third Party Auditor**: It is used in the situation where the user does not have required resources to avail the services of the cloud. It is an optional entity.

2.1. Cloud services

Services provided by the cloud are built upon three technologies models which are Platform as a service, Software as a service and Infrastructure as a service. These services are provided to users by their demand and requirement of their applications [18], [19]. Cloud Service Providers frame three distinct layers to implement different technologies in cloud system which are:

2.1.1. **Infrastructure as a service**: This level is subjected to do the task of management and storage of cloud resources. Most often cloud operates on virtual resources, and thus users can have access to a variety of virtual resources like servers, hardware, software, etc. that are provided to users to fulfill application’s requirement [33].

2.1.2. **Platform as a service**: The next level provides the platform to develop software and applications. It facilitates management and deployment of user’s application. It includes software and hardware tools provided by service providers [34]. It also accommodates application frameworks to support Software as a Service.

2.1.3. **Software as a service**: This level provides the most superior service which enables the cloud users to collaborate with the application. There is no need to install hardware and software resources at user’s place [35]. It is end user’s application where do not have to focus on management of infrastructure and maintenance of service.

2.2 Cloud security issues

Despite these advantages, we have a plenty of security issues related to cloud services due to which many organizations and users are unwilling to take cloud benefits. Some of the security issues are described by [28].

2.2.1. **Data breaches**: It occurs when the sensitive and confidential information of users stored in the cloud is stolen, viewed and thus exposed to unauthorized entities [36]. It includes trade secrets, personal health information, etc.
2.2.2. **Account hijacking**: It is the process where the hacker hijacks the login information of the user and use it for doing some unauthorized or malicious activities on remotely stored cloud data of the user [30].

2.2.3. **Insider attacks (threat)**: In this authorized employee of an organization misuses his granted privileges to get access to user’s sensitive information like account details, financial forms, etc. Most of the organization do not focus much towards this attack because their primary focus is towards external attacks [31].

2.2.4. **Malware injection**: It happens when the cloud services are embedded with code or scripts that perform like “legitimate instance” and operate on cloud server as SaaS [29]. It appears to do the logical operation, but actually, it facilitates hackers to eavesdrop and steal the sensitive information.

2.2.5. **Denial of service attack**: In this attack, the aim of the attacker is to flood the system, network, and services so that authorized users are not able to use them [37].

2.2.6. **Data loss**: Data Loss can occur through a malicious attack and natural disaster. It can also happen due to lack of recovery plan and improper handling and management of cloud data [32].

2.2.7. **Insecure APIs**: Application Programming Interfaces are used to customize the features of services provided by cloud according to the business needs [27]. With the growth in its infrastructure, security risk also increases. Insecurity in the API lies in the communication which occurs between applications.

### 2.3. Cloud security requirements

Although the services provided by the cloud are regularly being improved, still there is a great need for protection of data stored in the cloud. For this, the most important requirement is to build up trust between the user and service provider. The cloud infrastructure must be capable enough to implement the appropriate security measures at its premises.

To protect cloud data, following security measures need to be implemented [17]:

2.3.1. **Authentication**: This technique helps the communicating entities to prove its identity and assures authentic communication [21]. This service also guarantees that no other unauthorized entity can masquerade itself as authorized entity to take undue advantage of ongoing communication.

2.3.2. **Access control**: It is the process of imposing the restriction to access systems and applications according to the level of security requirements [21]. Authentication and identification of entity must be carried out to give access rights to the entity.

2.3.3. **Confidentiality**: Unauthorized exposure of information must be protected to maintain the confidentiality of sensitive cloud data [21]. The attacker is not allowed to look at frequency, length and other attributes of traffic flowing through the network.
2.3.4. **Integrity:** This service assures the correctness and validity of data being transmitted through the network. The received data must be free from duplication, modification, and reordering [21]. Only authorized users can make changes to it.

2.3.5. **Availability:** This service assures that information is available to authorized users whenever required [21]. To maintain it, offsite backup should be done regularly, and the systems must be protected against Denial of Service attacks.

2.3.6. **Non-Repudiation:** This service provides the proof that the alleged sender and receiver have sent and received the information respectively [21]. For it, accurate, traceable records must be maintained.

### 3. LITERATURE SURVEY

Song et al. [20] had proposed an idea where each word of the file is encrypted separately. But this technique resulted in higher cost as it required the word to word scanning of the documents. So this scheme was not efficient. They suggested a sequential scan that could be executed with or without an index. When the documents in the dataset are large, then the index-based scheme is preferred because it gives faster search results. But this system causes trouble in the situation where storage and updating of records are needed.

To make the searching process more user-friendly, Wang et al. [22] suggested a secure ranked keyword search method in which the matching files are returned in a ranked order depending on certain relevant standards like keyword frequency, etc. Wang et al. [22] advised a searchable symmetric encryption (SSE) and demonstrated how it was not efficient. Later they designed an order-preserving symmetric encryption scheme to deliver improved security in contrast to SSE system, together with the advantage of ranking results.

To get secure ranked search in encrypted cloud data Wang et al. [23] later carried on to put forward the concept of an encrypted invert index. It aimed at calculating the relevance score between query and documents. Depending upon the relevance scored calculated, documents are ranked so that users can fetch most relevant n results. It was noticed that due to lack of ranking mechanism, a lot of user time is wasted on searching for desired information from an enormous amount of records. So Li et al. [12] introduced and applied the concept of order-preserving techniques for faster retrieval of files.

Boneh et al. [2] proposed the idea of public key encryption system where keyword searching is done on encrypted data. This approach employs public key to store information in the cloud and uses private key for searching process.

Ballard et al. [1] devised conjunctive keyword searching techniques to enhance the search procedure. But these methods incurred significant overhead expenses in communication due to sharing the secret and increased computational costs due to bilinear mapping.
Pang et al. [16] present a privacy-preserving, similarity based text retrieval scheme where the search results are hidden from the unauthorized entities. Also, the server is unable to reconstruct the term composition of documents and queries performed. They employed similarity measure of “coordinate matching” organized as multi-keyword semantics. It uses “inner product similarity” to quantitatively evaluate the similarity measure. But there were two shortcomings of this scheme. First of all, it requires the reorganization of static dictionary each time with the entry of the new keyword. When the size of the collection of records grows exponentially then accordingly the time for search also increases exponentially.

Cao et al. [3] proposed an advanced scheme in which the number of computations was reduced with the increase in the size of keyword dictionary. Access frequencies of keywords were also taken into consideration. But this scheme was not user-friendly as it doesn’t have the features of semantics and fuzzy keywords.

To make the searching process more user interactive, fuzzy keyword searching were introduced over encrypted cloud data. Initially, the concept of fuzzy keyword searching for encrypted cloud data was given by Li et al. [11]. This technique attempted to make the search procedure user interactive. It states that the search system used in this method can give the accurate result even if the keyword is slightly misspelled by the user. On the other hand, in traditional techniques, no result is found when there are minor errors in spelling of keywords entered, and hence it makes the user’s task very complicated. To handle this problem, Li et al. [11] implemented fuzzy keyword searching. Not only that, but it also focussed on preserving the privacy of keywords. If user spell incorrectly then edit distance is used by fuzzy keyword to calculate the closest matching keyword. To diminish the difficulty in storage and to handle the issues in representation, they developed keyword dictionary. They demonstrated that their work was proficient in maintaining the privacy and security employing detailed security analysis. It also showed the utility of fuzzy keyword searching technique.

Khan et al. [8] tried to ameliorate the previous works in this field by adding the ranking functionality together with multi-keyword searching over encrypted cloud data and thus bettering the user search experience. This technique considered the relevance of results by some matched keywords. But this technique did not rank the results internally and also the synonym searching was not taken into consideration. And therefore the searching time was increased intensely.

Chai and Gong [4] presented a verifiable search technique in which they had verified the efficiency as well as integrity and accuracy of results.

Wang et al. [24] supported a fuzzy keyword searching method which considered verification based on VSSE (Verifiable symmetric searchable encryption). But this scheme neglected to rank results.

Fu et al. [5] further suggested a synonym based multi keyword search system in an encrypted cloud data. It might be possible that user forgets the exact keyword. Then the user can do searching by similar meaning words.
4. ALGORITHMS AND TECHNIQUES USED

4.1. Fuzzy keyword searching using wildcard

Fuzzy searching is a searching procedure in which matching is done in such a way the correct result is returned even if the entered word is incomplete or slightly misspelled. Fuzzy searching is performed on keywords which provide the functionality of downloading files.

The wildcard is a character which can represent more than one other character, and it is used to maximize search results [26]. Wildcard technique is an interactive searching method for term or query. The wild card is utilized to denote edit distance. The edit distance ed (w1, w2) is the number of operations needed to transform one word to another word among two words w1 and w2 [15], [21]. The edit distance has three operations:

1. Substitution: In a word, change one character to another character.
2. Deletion: Deletes one character from a word
3. Insertion: Adds one character to a word.

In this method, if the operation is performed in the same position then all alternatives is checked. For example, if the pre-edit distance is set to 1 for the word CASTLE then $S_{CASTLE, 1} = \{CASTLE, *CASTLE, *ASTLE, C*ASTLE, C*STLE, CASTL*E, CASTL*, CASTLE*\}$. Here total number of variants is $13 + 1$, rather than $13 \times 26 + 1$.

4.2 Algorithms used for Encryption and Decryption process

We have used two types of encryption algorithm to carry out the process of uploading and downloading of files. AES has been used as symmetric encryption, and RSA has been used as asymmetric encryption.

4.2.1. Advanced Encryption Standard (AES) algorithm

We are considering AES for encryption purpose because it is seen as most secure encryption algorithm till now [7], [10]. It is also faster in hardware and software implementation as compared to DES and RSA [14].

AES is symmetric encryption algorithm since the same key is used for encryption as well as decryption. In this algorithm input data is processed in the form of blocks of size 128 bits. It supports three distinct key sizes: 128, 192 and 256 bits which depend on the number of rounds. In most of the cases, the key size of 128 bits is chosen [21]. So we have opted key size of 128 bit which has ten processing rounds for encryption.

AES is a non-Feistel structure in which the parallel processing of input block is done using permutation and substitution operations in each round [21]. The decryption process is identical to encryption, but the round keys are applied in inverse order.

4.2.2. RSA Algorithm
Ron Rivest, Adi Shamir, and Leonard Adleman developed an asymmetric encryption algorithm which used two different keys for encryption and decryption. One is named as the public key and the other as the private key [21]. As the name indicate, the public key is known to each user in the network and the private key is kept secret. For authentication, the message is encrypted with the private key, and for confidentiality, the message is encrypted with the public key. For decryption of produced ciphertext, opposite keys are used. Both the keys are mathematically linked to one another [21].

Out of the two keys, one is used to encrypt, and the other is used to decrypt the message. RSA is considered as most extensively used asymmetric algorithm as it offers confidentiality, integrity, authentication and non-repudiation services to data storage and communication [38].

RSA is secure because the multiplication of prime number used is simple, but estimation of the original prime number from the total is hard to guess as it would take a lot of time [38].

5. PROBLEM FORMULATION

We deal with cloud user, cloud server in a cloud network. Given a set of n encrypted files \( F = (F_1, F_2, F_3....F_n) \) and a set of keywords \( K = (K_1, K_2, K_3....K_n) \) saved in cloud server then authorized users are allowed to search their desired files over the encrypted data \( F \) in the cloud with the help of keyword \( K \).

We consider legal authorization between the user and data owner in the cloud. The user inputs his request to cloud system to get his desired file of interest. Files are stored on the server based on file Id and each of them the is mapped to keywords linked with them. The cloud server associates each keyword search request with the relevant files.

The following rules are followed for retrieving the needed files based on fuzzy keyword searching:

1. If the entered keyword exactly matches with the saved keyword, then the server must return file corresponding to given keyword.
2. If there exist any spelling error or format disparity in the entered keyword, then the server must return the nearest probable result by pre-specified resemblance semantics.

When the keyword based searching is carried out in the cloud, then it might be possible that any unauthorized entity attempts to impersonate as the legitimate server and try to gain access to some sensitive information regardless of secure AES encryption. For this reason, the searching technique should be performed in such a way that the user can authenticate the server before retrieving files. To achieve this, we have added RSA algorithm along with AES algorithm for secure file retrieval from the cloud server.

This paper offers an efficient solution for cloud users to verify the authenticity of server and securely store and retrieve their needed data to and from the cloud.
The aim of this work is to:
1. To provide strong authentication to the server so that users can download files securely.
2. To devise a search procedure based on the developed fuzzy keyword searching technique.
3. To provide the security to above-designed information retrieval system by adding RSA algorithm to AES algorithm.

6. PROPOSED WORK
Authorized users are added by the organization who need access to their data. The user enters the query of file retrieval by inserting the keyword. Compare the entered keyword with the stored keyword. If it matches, then the requested file is sent back to the user after decryption. We have proposed a strong server authentication system over the stored encrypted data. Techniques implemented up to now has focussed on fuzzy keyword searching on encrypted cloud data. In this work, we have added dual encryption by enforcing the combination of the symmetric as well asymmetric encryption algorithm.

7. IMPLEMENTATION
The algorithm would be devised and implemented on JDK environment. We shall then calculate the total time required to upload and then download the file. Hence we shall combining the encryption and decryption time for the whole cycle.
Fuzzy Keyword searching is performed using the AES encryption but our proposed keyword searching scheme implements it using the combination of AES and RSA algorithm which is named as the dual encryption algorithm. Our proposed scheme is combining RSA and AES so that authentication is also provided along with security. Although our scheme is slower than traditional techniques, it can work better in situations where authentication is required.
The user uploads his files along with the corresponding keyword so that he can perform the fuzzy keyword search and later download the file using the same keyword in a secure fashion. The user can verify that his file is downloaded from an authentic server.

8. CONCLUSION
This paper suggests an efficient and secure keyword based searching scheme where the user can store his files in a secure manner. To make the searching procedure user interactive, fuzzy keyword searching is introduced using the wildcard technique. The cloud server is semi-trusted where the user’s data is stored in encrypted form to prevent server and unauthorized party from learning any information regarding the
stored user’s data. In the proposed approach strong security shall be provided with server’s private key. In this way, the cloud system becomes more resistant to different security attacks performed by unauthorized entities who try to disclose the sensitive user’s information for their benefits.

REFERENCES


Secure Keyword Search Using Dual Encryption in Cloud: An Approach


[31] https://www.techopedia.com/definition/26217/insider-attack


[33] http://searchcloudcomputing.techtarget.com/definition/Infrastructure-as-a-Service-IaaS

[34] http://searchcloudcomputing.techtarget.com/definition/Platform-as-a-Service-PaaS


[38] http://searchsecurity.techtarget.com/definition/RSA