** P.E.S. COLLEGE OF ENGINEERING,**

**MANDYA, 571401**

**(An Autonomous Institution under Visvesvaraya Technological**

**University, Belgaum)**

A Project Phase-1 Report on

**Fragmentation and Encryption of Files to**

**Avoid Data Breaching in Cloud**

In partial fulfillment of the requirement for the award of the Degree

In

**COMPUTER SCIENCE AND ENGINEERING**



Submitted by

**BHUVAN JAIN S M [USN: 4PS18CS022] B IKSHWAK [USN: 4PS18CS016] SUHAS N KASHYAP [USN: 4PS18CS107] DEEPAK C [USN: 4PS18CS028]**

Under the guidance of

**Mrs. SHWETHA M K**

Asst. Professor, Dept of CS&E, P.E.S.C.E, Mandya.

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**P.E.S. College of Engineering, Mandya. 2020-2021**

**P.E.S COLLEGE OF ENGINEERING, MANDYA**

**(An Autonomous Institute under Visvesvaraya Technological University, Belagavi)**

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**



CERTIFICATE

This is to certify that the project entitled “**Fragmentation and Encryption of Files to**

**Avoid Data Breaching in Cloud**” is carried out by **Bhuvan Jain S M [4PS18CS022], B Ikshwak [4PS18CS016], Suhas N Kashyap [4PS18CS107], Deepak C [4PS18CS028]**, bonafide student of

P.E.S College of Engineering, Mandya, in partial fulfilment for the award of the Degree of Bachelor of Engineering in Computer Science and Engineering of Visvesvaraya Technological University, Belagavi during the year 2020-2021. It is certified that all corrections/suggestions indicated in internal presentation have been incorporated in the final report. The project report has been approved as it satisfies the academic requirements in respect of internship prescribed for the said degree.

**Mrs. Shwetha M.K Dr. Nagarathna**

**Assistant Professor, Professor & Head,**

**Dept of CSE Dept of CSE**

|  |  |  |
| --- | --- | --- |
| **Sl.No** | **Name** | **Signature** |
| **1** | **Dr. Minavathi**  Professor and Dean (Research) |  |
| **2** | **S.K Shivshankar**  Assistant Professor |  |
| **3** | **Deepika**  Assistant Professor |  |

**Table of Contents**

[1. Chapter 1 1](#_Toc106624609)

[1.1. Introduction 1](#_Toc106624610)

[1.2. Aim 3](#_Toc106624611)

[1.3. Objective 3](#_Toc106624612)

[1.4. Existing System 4](#_Toc106624613)

[1.5. Proposed System 5](#_Toc106624614)

[2. Chapter 2 7](#_Toc106624615)

[2.1. Literature Survey 7](#_Toc106624616)

[3. Chapter 3 9](#_Toc106624617)

[3.1. Requirements 9](#_Toc106624618)

[3.2. Hardware and Software Requirement 11](#_Toc106624619)

[References 12](#_Toc106624620)

**Fragmentation and Encryption of Files to Avoid Data Breaching in Cloud**

**ABSTRACT**

Cloud is used in various fields like industry, military, college, etc. for various services and storage of huge amount of data. Data stored in this cloud can be accessed or retrieved on the users request without direct access to the server computer. But the major concern regarding storage of data online (i.e, cloud storage) is the Security. This Security concern can be solved using various ways, the most commonly used techniques are cryptography and steganography. But sometimes a single technique or algorithm alone cannot provide high-level security. So, we have introduced a new security mechanism that uses a combination of multiple cryptographic algorithms of symmetric key and steganography. This document is a broad survey of the different approach which is used for securely storing files, and sharing it over the network. This proposed scheme will also ensure the whole model to have confidentiality, integrity, and availability mechanisms to be implemented in it. In this proposed system 3DES (Triple Data Encryption Standard), RC6 (Rivest Cipher 6) and AES (Advanced Encryption Standard) algorithms are used to provide security to data. File during encryption is split into three parts. These individual parts of the file will be encrypted using different encryption algorithm simultaneously with the help of multithreading technique.

**Department of CS&E, PESCE, Mandya**

# Chapter 1

### Introduction

Cloud computing is a broad and rapidly developing field. It may have different meanings for different individuals, but the high availability of data at any time and in any location is a common feature that connects individuals together. Cloud computing not only reduces the role of local computers but also makes computing more integrated. The word 'Cloud' computing comes from two words, that is Cloud which refers to the internet and 'Computing' which means technology based on computers. Here, Internet is storage on warehouse where the virtualized resources are stored which are provided as services. From building through initial concepts to the actual deployment, cloud computing has been expanding. By adapting to the cloud computing techniques, the growth in business development will be more efficient and data more secure. The old-style approach of storing documents was to write them in MS Word but that might be substituted by Cloud Computing. It is a more effective way of doing that task as the user can just log into his account and use the Google Document Service provided by Google. On the other hand, storing the data on cloud can make it more prone to threats and attacks. Thus, the concern of security and privacy of data is of utmost importance. Cloud computing is utilized in a variety of sectors, including industry, military, and higher education, for a variety of services and storing of large amounts of data. Without direct connection to the server computer, data stored in this cloud can be viewed or retrieved at the user's request. The cloud storage is a storage space available to store data on remote servers which is placed in the data centers that are located anywhere in the world and maintained by the third party. When the user wants to retrieve the information, they access the data server through a web- based interface. In Cloud computing files and software are not fully contained on the user’s application and Program are residing in provider premises. The cloud provider can solve this problem by encryption the files by using encryption algorithm. However, security is the primary concern when it comes to data storage on the cloud. This security issue can be addressed in a number of ways, the most common of which are cryptography and steganography. This document presents a file security model to provide an efficient solution for the basic problem of security in cloud environment. In this model, encryption is done on files by file splitting and applying a set of algorithms on each segmented files to ensure security of the file. A computer cloud is a target-rich environment for malicious individuals and criminal organizations. It is thus no surprise that security is a major concern for existing users and for potential new users of cloud computing services.

There are multiple ways to look at the security risks for cloud computing. A recent paper identifies two broad classes of risk: traditional security threats and threats related to third-party data control.

1. Traditional threats are those experienced for some time by any system connected to the Internet, but with some cloud-specific twists.
2. Third-party control generates a spectrum of concerns caused by the lack of transparency and limited user control. For example, a cloud provider may subcontract some resources from a third party whose level of trust is questionable. There are examples when subcontractors failed to maintain the customer data. There are also examples when the third party was not a subcontractor but a hardware supplier and the loss of data was caused by poor-quality storage devices

In a word, to protect cloud data from leakage at storage layer, this document presents Cloud Secure Storage Mechanism, a Cloud Secure Storage Mechanism. Cloud Secure Storage Mechanism combines data dispersion with data encryption, so that large-scale cloud data and keys would be stored in chunked cipher texts. On this basis, user password and secret sharing are introduced to further protect keys security. It majorly focuses on:

1. Data Secure Storage: In order to prevent data leakage and increase the difficulty of attack, this document presents a method combining data distribution and data encryption to improve data storage security.
2. Hierarchical Key Management: To protect the key and prevent the attacker from using the key to recover the data, this document introduces secret sharing and key hierarchy derivation algorithm in combination with user password to enhance key security.
3. Experimental Evaluation and Analysis: The security analysis and experimental results show that Cloud Secure Storage Mechanism can effectively guarantee the security of data storage, and the increased performance cost is acceptable to users.

Mobile devices, such as smart-phones and tablets, record static images, as well as movies and have limited local storage capacity, so they transfer the data to cloud storage systems. Sensors, surveillance cameras, and digital medical imaging devices generate data at a high rate and dump it onto storage systems accessible via the Internet. Online digital libraries, eBooks, and digital media, along with reference data, add to the demand for massive amounts of storage.

### Aim

Based on the survey it was identified that secure file storage and sharing would not only require confidentiality but also authentication and integrity. To overcome these drawbacks an architecture is proposed which tries to provide a complete solution for securely storing the files.

In this document the use of multiple encryption technique outlines the importance of data security and privacy protection by using effective encryption methods to increase data security.

### Objective

* The main objective is to prevent the unauthorized access by unauthorized users from accessing files that are present in cloud.
* Providing solutions to data theft insecurities on cloud.
* Here file encryption can be achieved that are stored in cloud using AES-128, AES-192,

AES-256 advanced algorithms.

* And to achieve decryption of file by applying same algorithm in reverse order.

### Existing System

Data Security Issues are the main issue in the existing system. Due to the multi-tenant characteristics of the cloud, the previous security mechanisms are no longer suitable for data in the cloud. Data of the cloud platform have no fixed infrastructure and security boundaries. Due to the openness of the cloud and sharing virtualized resources by multitenant, user data may be accessed by other unauthorized users.

##### Limitations:

* + Without any security to the file unauthorized user can access the file.
  + Usage of brute force attacks makes it very easy to access the files.
  + Although a number of techniques have been projected previously but it’s far from the security concern.
  + MD5 (Message Digest 5) hashing is used but it’s more vulnerable to security breach.

More work is required in the area of cloud computing and the gaps in data security which makes it more vulnerable can be filled by making more effective techniques though a number of techniques have been projected previously by researchers for data protection as well as to attain the maximum level of data security in the cloud but more elaborated work on them as well as the upcoming techniques are needed.

### Proposed System

A method for securely storing files in the cloud utilizing a hybrid cryptography algorithm is provided in the suggested system. The user can safely save files in the cloud storage with this method since the files are stored in encrypted form in the cloud and only the authorized user has access to their files.

The user must first register in order to use the services. During the registration process, you will be asked to input information such as your name, username, password, email address, and phone number. The key will be generated randomly, which will be used for encryption and decryption. This keys, however, will not be saved in the database. Instead, it will be hidden in an image that will be used as the user's profile picture, using the steganography technique.

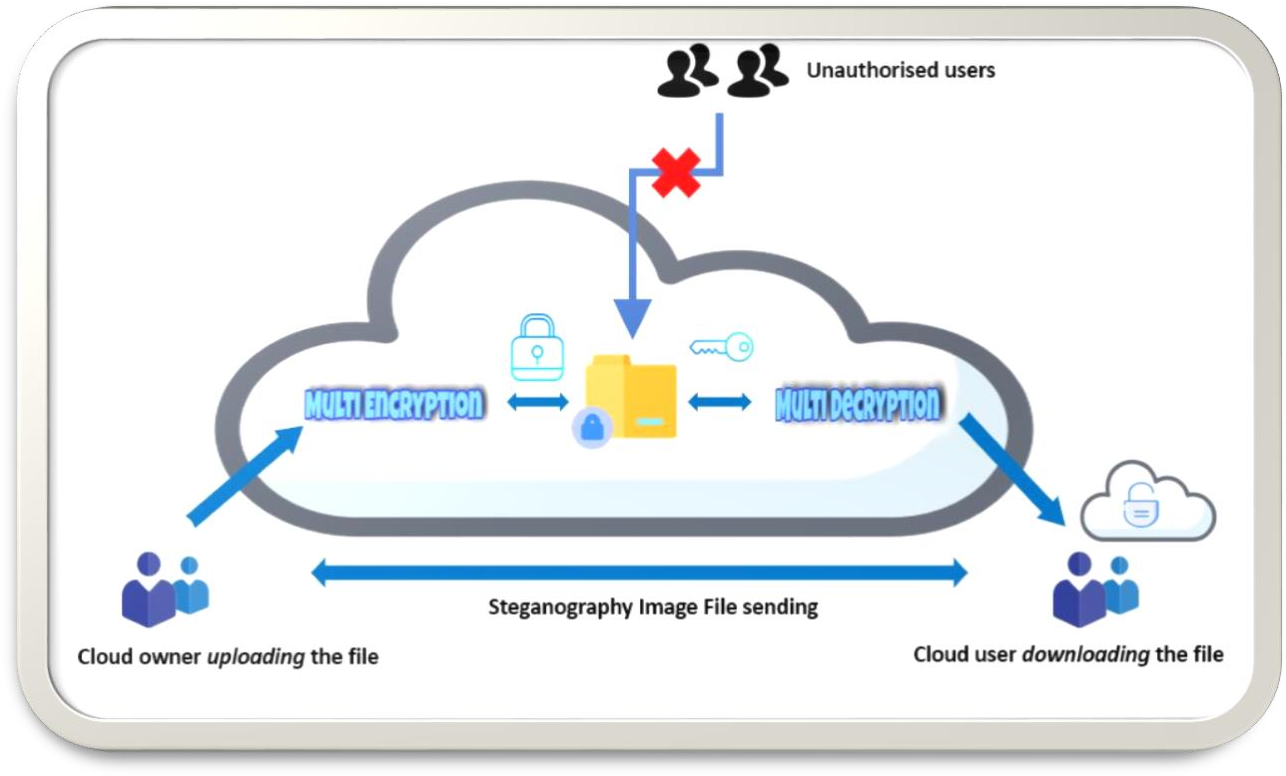


Figure 4.0 Working of the Proposed System

As shown in the above figure 4.0 it gives the brief picture on how the encryption and decryption of the files works.The cloud owner uploads the file which undergoes multiple encryptions before it is sent to the cloud storage and this file which is present in the cloud can be accessed by the user by requesting key to the cloud owner.

If the cloud owner grants the key to the user, then the file undergoes multiple decryptions and the final original file is securely accessed by the cloud user.

In case, if an unauthorized user tries to access the files from cloud storage, he will not get the file, in the worst case if he gets access to it without the proper key by hacking, the file will be in the encrypted format.

##### Uploading a File on Cloud

* When the owner uploads a file on the cloud, the file will be split into 3 parts.
* These all parts of file will be encrypted using cryptographic algorithms. Every part will use a different encryption algorithm namely AES-128, AES-192, AES-256. The key to these algorithms will be retrieved from the steganographic image.
* After the split encryption all three parts of the files will be reassembled and stored in AWS S3 Bucket.

##### Downloading a File from the Cloud

* When the user requests a file to download, first the encrypted file is fetched from AWS S3 Bucket and split into 3 parts.
* These 3 parts will be decrypted by the same set of algorithms using the keys sent by the owner when requested.
* These parts will be reassembled to form a fully decrypted file (original file).
* Finally, the decrypted file will be downloaded to the user’s device.

# Chapter 2

### Literature Survey

According to the United States National Institute of Standards and Technology( NIST) Computing in cloud has been defined as, “ A model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.”

Following are some papers been referred to:

1. Title- Secure file storage in cloud computing using a hybrid cryptography algorithm.

Author - Punam V. Maitri, Aruna Verma, Year – 2016

Description – The paper focuses on how files are securely stored on a cloud platform. Also, it discusses the problem of using only a single algorithm to encrypt the file and how ineffective it will be on the cloud. This paper splits the file into blocks and each block is encrypted using AES, Blowfish, RC6 algorithm. The key information about which file uses which algorithm is sent to the receiver using steganography modern approach to file system integrity checking. The security of the information on cloud relies on classical principles of availability, integrity and confidentiality though, applied to distributed virtualize and dynamic architecture says the author.

Limitation – Using Blowfish encryption-decryption efficiency is less [1]

1. Title-Secure cloud auditing over encrypted data.

Author-Shaikh, S., & Vora , D, Year- (2016)

Description – This is Multi Encryption & Decryption algorithm where authorized user can access the data. No other intruder gets the data, he must have to decrypt the data at each level which is a very difficult task without a valid key. The growing competition in this next generation platform is an upcoming concern in the area of security. Such storage in a third party cloud implies increased vulnerability for security and confidentiality. While practices like data masking, authentication, encryption and integrity checking have already come up in the markets, yet end-to-end security still remains an unanswered issue. As cloud computing is at a very early stage, Mr. Shaikh states that future technologies and modern designs pave a way for further research.

Limitation – It is time-consuming as multiple encryption and decryption take place. Secure data sharing using cryptography in a cloud environment [2]

1. Title-Secure File Storage Using Hybrid Cryptography

Author- Ronak Karani, TejasChoudhari, Anindita Bhajan, Madhu Nashipudimath (2020) Description- This paper focuses on providing the facility to securely store and share the data in a group using cloud technology for storage. This paper overcomes the security tradeoff and improves the performance of data transmission and increases security. But concerning some business-critical requests, the administrations, specifically huge enterprises, still would not transfer files to cloud because of secrecy and data protection issues in cloud.

Limitation - MD5 hashes are no longer a secure way for encryption. [3]

1. Title- Enhancing Distributed Data Storage Security for Cloud Computing AES Algorithm. Author - Tulip Dutta, Amarjyoti Pathak, Year – 2016

Description – The proposed security mechanisms will prevent confidential data from being misused making the system more reliable. High speed: The proposed method will make encryption and decryption with proper keys much faster than usual. This paper discusses how a secret key can be shared with other users to whom access needs to be given. The problem with using a single key to encrypt all data and using different keys for different files. The solution to this is using key aggregation. In key aggregation, different data files are encrypted with different keys and then for decryption, a single aggregated key is used. Satisfying confidentiality, integrity, and authenticity. [7]

Limitation – Using AES Algorithm nowadays compromises the data stored on cloud and having a single algorithm does not completely secure the data as it will be vulnerable, attackers can easily crack the key. Hence when it comes to efficiency it is much lesser then our proposed paper.

1. Title- Secure algorithm for cloud computing and its applications.

Author- Bhandari, A., Gupta, A., & Das, D.-2016

Description – This paper describes an approach to the integrity of files and restoring the files if any problem is arising in the future. This proposed course uses a pattern of each protected file to determine its modification. Methods used for pattern generation are cryptographic hash functions. This system uses a database that stores the names of all files that are to be protected and their hash codes. To check the integrity of the file the hash code of the file is produced and checked with one in the database. After the file is verified then only access is granted. [8]

Limitation- Databases can degrade if they go through a **large number of collisions.**

# Chapter 3

### Requirements

A System Requirements Specification (SRS) (also known as a Software Requirements Specification) is a document or set of documentation that describes the features and behavior of a system or software application. It includes a variety of elements that attempts to define the intended functionality required by the customer to satisfy their different users. In addition to specifying how the system should behave, the specification also defines at a high-level the main business process that will be supported, what simplifying assumptions have been made and what key performance parameters will need to be met by the system. The SRS is often referred to as the "parent" document because all subsequent project management documents, such as design specifications, statements of work, software architecture specifications, testing and validation plans, and documentation plans, are related to it.

System requirement plays an important part in the analysis of the system. Only when the requirement specification is properly given, it is possible to design the appropriate system, which will fit into a pre-defined and required environment.

The requirement specification for any system can be broadly stated as given below:

* The system should be able to interface with existing system.
* The system should be better and accurate than existing system.
* The system should be interactive so that its user should easily understand the operation.
* The system should show the appropriate message when the user finishes the operations.

Also, when there is any error condition, suitable message should be displayed. The requirement analysis phase is for understanding the existing system, which is usually a simple activity. Recognition of Phony Reviews in E-commerce Platform Department of CS&E, PESCE, Mandya Page 13 The goal of this activity is to understand the requirements of the good system to be the current mode of operation are basic requirements available for the analysis. The process of obtaining answers to questions that might arise in an analyst’s mind continues until the analyst feels that enough information has been obtained. It is important to note that an SRS contains functional and non-functional requirements only.

Functional Requirement:

Functional requirements in an [SRS document (software requirements specification)](https://winatalent.com/blog/2020/05/srs-document-the-what-the-why-and-the-how/) indicate what a software system must do and how it must function; they are product features that focus on user needs. These are the requirements that the user specifically demands as basic facilities that the system should offer and this segment normally comprises a hierarchical arrangement of requirements, with the functional/business requirements at the uppermost level and the detailed system requirements are listed as their child items. It also defines a function of a software system and how the system must behave when presented with specific inputs or conditions. In software engineering and systems engineering, a Functional Requirement can range from the high-level abstract statement of the sender’s necessity to detailed mathematical functional requirement specifications. These may include calculations, data manipulation and processing and other specific functionality. There are several methods to write functional requirements, but the most common method is by constructing user stories.

Benefits of Functional Requirement are as follows:

* Helps us to check whether the application is providing all the functionalities that were mentioned in the functional requirement of that application
* A functional requirement document helps you to define the functionality of a system or one of its subsystems.
* Functional requirements along with requirement analysis help identify missing requirements. They help clearly define the expected system service and behavior.
* Errors caught in the Functional requirement gathering stage are the cheapest to fix.
* Support user goals, tasks, or activities

Non-Functional Requirement:

A non-functional requirement is a specification that describes the system’s operation capabilities and constraints that enhance its functionality. These may be speed, security, reliability, etc., and also defining the quality attribute of a software system, they specify the criteria that can be used to judge the operation of a system rather than specific behaviors.

Benefits of Non Functional Requirement are as follows:

* The non-functional requirements ensure the software system follow legal and compliance rules.
* They ensure the reliability, availability, and performance of the software system
* They ensure good user experience and ease of operating the software.
* They help in formulating security policy of the software system.

### Hardware and Software Requirement

The software requirement specification is produced at the culmination of the analysis task. The function and performance allocated to software as part of system engineering are refined by establishing a complete information description, a detailed function description, representation of system behavior, an indication of performance requirement and design constraints appropriate validation criteria, and other information pertinent to requirement. The introduction to software requirements specification states the goals and objectives of the software, describing it in the context of computer-based system. The Information Description provides a detailed description of the problem that the software must solve. Information content, flow and structure are documented. A description of each function required to solve the problem is presented in the Functional Description. Validation Criteria is probably the most important and ironically the most often neglected section of the software requirement specification. Software requirement specification can be used for different purpose.

###### Hardware Requirements:

* + Operating System: Windows, Linux, MacOS
  + Memory (RAM): 4GB or more
  + Disk Space: 100GB

###### Software Requirements:

* + NodeJS installed on your machine
    - NodeJS is used for Server-Side development
    - Cryptographic Algorithms can be imported to our project using Node Package - crypto
  + XAMPP installed on your machine
    - This is used for our database purposes - MySQL

Web Browser and Server

## References

1. Maitri, P. V., & Verma, A. (2016). Secure file storage in cloud computing using a hybrid cryptography algorithm. 2016 International Conference on Wireless Communications, Signal Processing and Networking.
2. Shaikh, S., & Vora, D. (2016). Secure cloud auditing over encrypted data. International Conference on Communication and Electronics Systems (ICCES).
3. Ronak Karani, TejasChoudhari, Anindita Bhajan, Madhu Nashipudimath (2020). Secure File Storage Using Hybrid Cryptography, International Journal of Innovative Research in Technology.
4. Heqing Song, Jifei Li, Haoteng Li (2021). A Cloud Secure Storage Mechanism Based on Data Dispersion and Encryption, IEEE Access Vol 9.
5. Peterson, Dawn Song. 2011. Remote data checking using provable data possession. ACM Transactionson Information and System Security (TISSEC) 14, 1, 12.
6. Jasleen K., S.Garg ,“Security in Cloud Computing using Hybrid of Algorithms ”, IJERJS, Vol 3, Issue 5, ISSN 2091-2730, September-October, 2015.
7. Shimbre, N., & Deshpande, P. (2015). Enhancing Distributed Data Storage Security for Cloud Computing Using TPA and AES Algorithm. 2015 International Conference on Computing Communication Control and Automation.
8. Bhandari, A., Gupta, A., & Das, D. (2016). Secure algorithm for cloud computing and its applications. 2016 6th International Conference - Cloud System and Big Data Engineering (Confluence)