JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Efficient and Secure Cloud Storage: Password-Protected Encryption Keys for Deduplication Systems

Pranav Patil[1], Prof. S. S. Bhosale[2], Niraj Mahajan[3], Rushikesh Patil[4]
Department of E&TC, SKNCOE, SPPU, Pune

[1] pranavpatil.13072001@gmail.com, [2] sonali.bhosale_skncoe@sinhgad.edu
[3] nirajnm.a01@gmail.com, [4] patilrushikesharun@gmail.com

Abstract **— Efficient management of files and storage has become crucial in today's world to prevent wastage of cloud storage resources. One commonly used technology for thispurpose is data deduplication, which eliminates duplicate copies of files on cloud servers and can significantly reducestorage requirements and save bandwidth. As a result, it can lead to substantial cost savings for cloud service users. Since data security is also a top priority, data is typically stored in encrypted form, and the use of personal encryption keys by data owners can prevent cloud service users from deduplicating their data.**

Keywords **—** ***Deduplication, Encryption, Decryption,Ciphertext, AES, MD5.***

I. INTRODUCTION

The utilization of computers, databases, mobile computing, and online applications has grown considerablyin recent times. This has led to a demand for more advanced and efficient data management systems. With an increase in data volume, storage space has become a significant concern. Consequently, cloud computing has become a preferred choice as it allows people to store their data remotely, which is less burdensome due to reduced storagelimitations, maintenance, and overhead costs. Responsibility re-encryption is a method of sharing accesskeys, enabling authorized users to access specific information without compromising sensitive data.

We avoid text and digital images in our work. For example, we have personal images on mobile devices, in today's world, various devices such as handhelds, desktops,and others are used to capture and store images. To ensurea high level of security, it is essential to store these imagessecurely with encryption. Textual data is also critical for modern users, and it should be stored on a cloud server thatis relevant to their work. Therefore, we employ suitable techniques to guarantee the secure storage of textual data

II. RELATED WORK

The research paper by S. Uthayashangar, J. Abhinaya, Harshini, and R. Jayavardhani titled "Image and Text Encrypted Data with Authorized Deduplication in Cloud" discusses the concept of de-duplication and explores three algorithms, namely the Levenshtein distance algorithm,fuzzy matching algorithm, and dice coefficient algorithm. To efficiently store a large amount of data while avoiding duplication of images and text, they employed a 2D cellular automata encryption method. In the paper "Secure Block-level Data Deduplication approach for Cloud Data Centers", Gulsayyar Ali, Dr. MianIlyas Ahmad, and Arslan Rafi focus on maximizing disk space savings and providing secure de-duplication mechanisms to maintain customer confidence in the system. The paper emphasizes the significance of information security and the need for a trustworthy system to safeguard customer data in cloud data centers.

As technology advances, the demand for storage in cloud data centers continues to increase. Data de-duplication is used to identify duplicate copies of files and reduce storageneeds. In "Secured and Reliable File Sharing System with Deduplication using Erasure Correction Code," Chippy Jacob and Rekha V. R emphasize the importance of efficient file storage and management in cloud computing.They propose a system that uses erasure correction code technology to ensure the security

and authenticity of encrypted files, even if part of the file is lost due to an attackon the server. Similarly, in "Private data deduplicationprotocols in cloud storage" by W.K. Ng, Y. Wen, and H. Zhu, a private data deduplication protocol is introduced andformalized. This protocol allows a client to prove ownership of private data to a server holding a summary string of the data without revealing any additionalinformation. The security of these protocols is formalized using a simulation-based framework for two-party computations, and a construction of private deduplication protocols based on standard cryptographic assumptions is presented and analyzed.

In the paper titled "Enhancing Efficiency and Security in Proof of Ownership for Deduplication" by R. Oi Pietro and A. Sorniotti [5], it is mentioned that the deduplication technique involves storing a single copy of a file that is identical to other files, which can be helpful for multiple users who want to store the same content. However, this technique introduces several security risks, and the paper focuses on addressing the most severe one. The paperdiscusses scenarios where an adversary, who only possesses a fraction of the original file or is partially colluding with a rightful owner, falsely claims to possess the entire file. The paper presents several contributions. Firstly, it introduces a new Proof of Ownership (POW) scheme that possesses all the features of the current leadingsolution, but with significantly less overhead. Secondly, the security of the proposed mechanisms relies on information theory and combinatorics, rather than computationalassumptions.

KEY CONCEPTS

A. Data De-duplication

Data deduplication is a technique used to eliminate duplicate copies of data, thereby reducing storage requirements. It can be performed either in real-time duringdata storage or as a post-storage process. Implementation of data deduplication can enhance storage utilization, resulting in cost savings by minimizing the amount of datastored to meet the storage capacity requirements. Additionally, it can be leveraged to optimize data transfersover a network, thereby reducing the amount of data transmitted.

Figure 1. Deduplication reduces the amount of stored data.

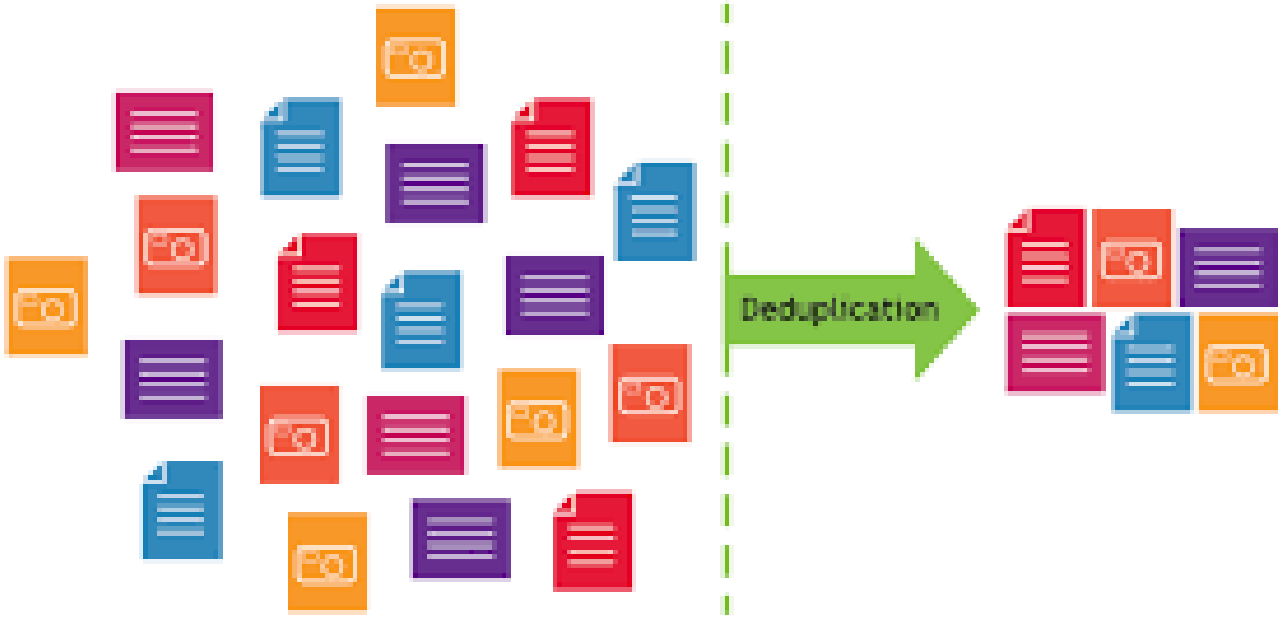

Fig. 1. Data Deduplication

B. Encryption and Decryption

Encryption is a technique used to secure information from unauthorized access by transforming it into an unreadable format called ciphertext. This process involves converting human- readable data into a form that appears random. Encryption relies on an encryption key, which is a set of numbers agreed upon by both the sender and the receiver of the message. Decrypting the encrypted data involves reversing the encryption process, and it requires a key or password for authorized access. Privacy is one of the main reasons for implementing an encryption-decryption system.

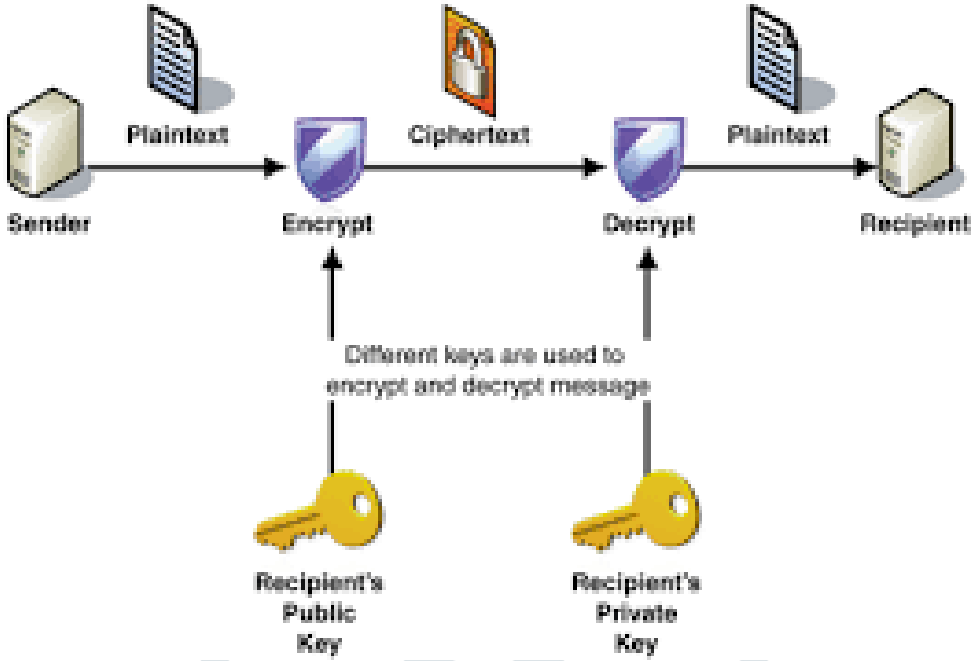

Fig 2. Sample Encryption-Decryption Process

III. METHODOLOGY

○ *AES Algorithm*

The US government chose the Advanced Encryption Standard (AES) as a symmetric block cipher to secure confidential information in various applications such as wireless security, processor security, data encryption, and SSL/TLS. Many organizations, including government agencies, non-profits, and businesses, rely on AES encryption to safeguard sensitive data. The AES encryptionalgorithm involves applying a set of modifications to the data stored in an array. The encryption process starts by placing the data into an array and then repeating the conversion multiple times.

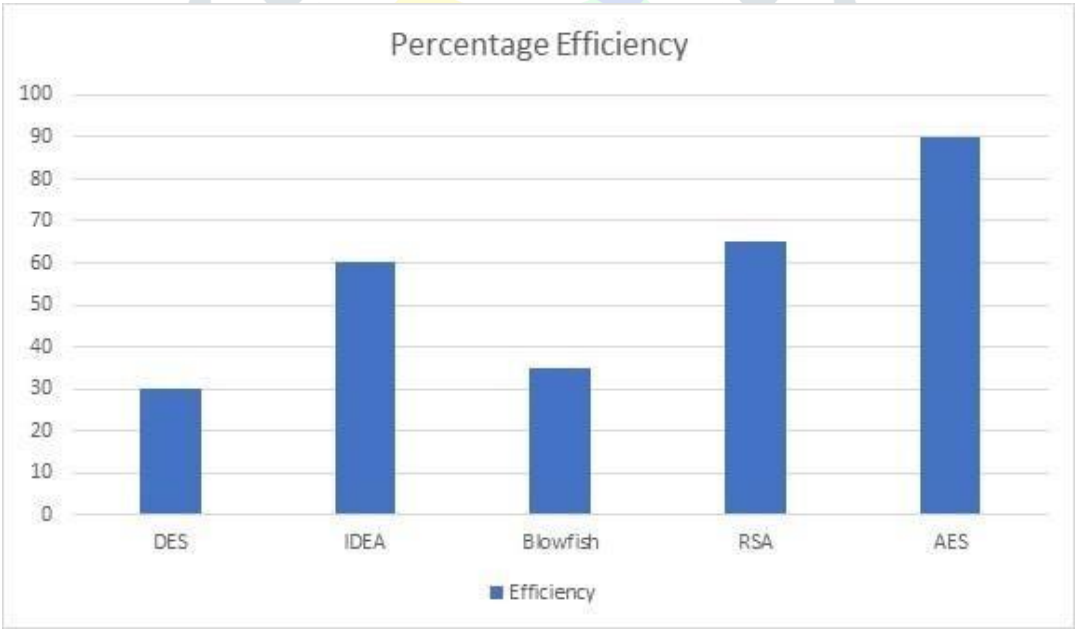

Fig 3. Efficiency rate

○ *MD5*

MD5 hashes are commonly referred to as "digital fingerprints" and serve as a distinct identifier for electronicinformation. They are frequently utilized to verify the authenticity of a document or to demonstrate that twodocuments have identical content. In real-world scenarios, MD5 is frequently utilized to verify the authenticity of data or to establish the likeness between two files. A MD5 digest is a 128-bit value that is typically represented as a 32-bit hexadecimal number. To generate this value, the binary content of a file is hashed using the MD5 algorithm.

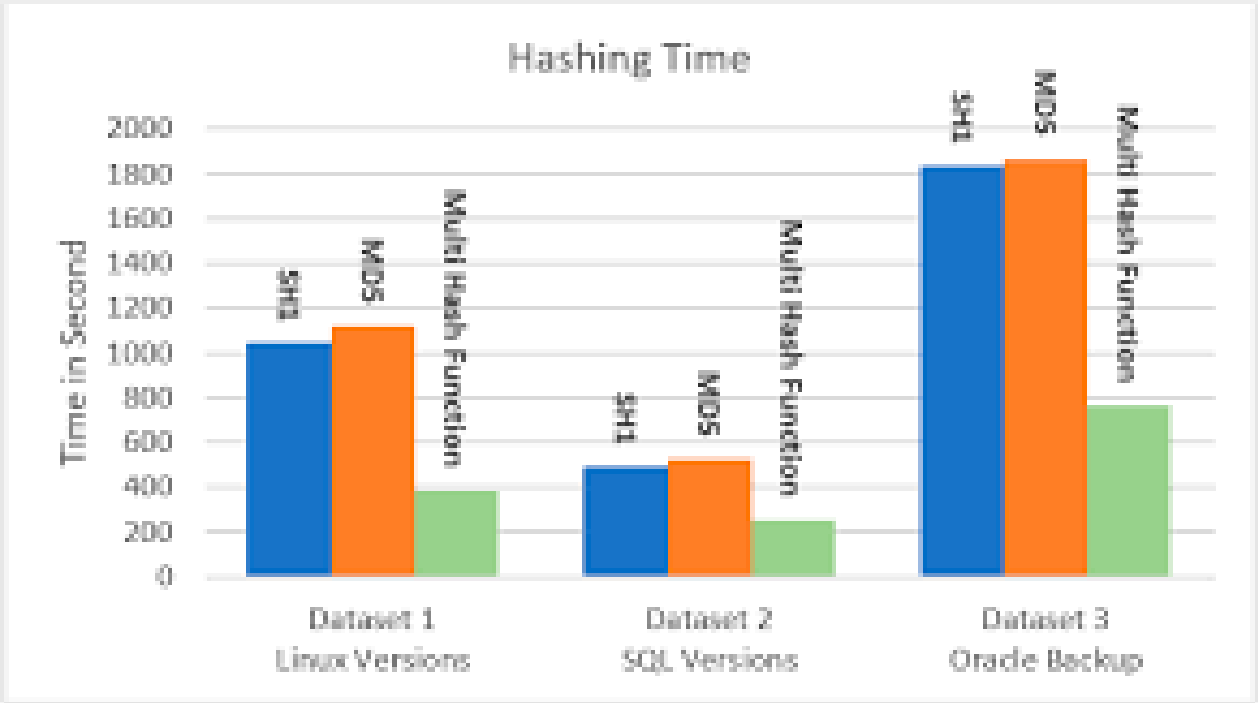

Fig 4. Difference between Hashing time of various algorithms

IV. SYSTEM ARCHITECTURE

To use the desktop application, it is necessary for the user to register beforehand, as access to the application is restricted to registered users only. Once the user completes the registration process, a unique user ID and password will be provided to access the system. The user can proceed to upload a text file or image in a suitable format after the registration process. Subsequently, the content of the uploaded file will beverified by the MD5 algorithm, also known as the "Message Digest Algorithm". If the content of the file matches with an existing file, the system will detect it as a duplicate and notify the user through a pop-up message on the screen. If the content of the uploaded file is found to be valid and not a duplicate, then it will be saved on the local cloud storage. To access an AES-encrypted file or data, the user is required to provide a unique key as a prerequisite for the subsequent procedures.

PARAMETERS-

- The performance of our model relies on certain factors including:
- The speed of MD5 algorithm for hashing.
- The speed of AES algorithm for encryption and decryption.The effectiveness of both AES and MD5, and the ability ofMD5 to detect duplicates. MD5, also known as Message Digest Algorithm, has a high efficiency for hashing speed. Then, AES is utilized for faster encryption and decryption,surpassing other algorithms in performance efficiency. These algorithms have a significant impact on the model's overall performance. MD5 primarily functions to identify duplicates within specified data, demonstrating its ability and capacity to perform this task effectively.

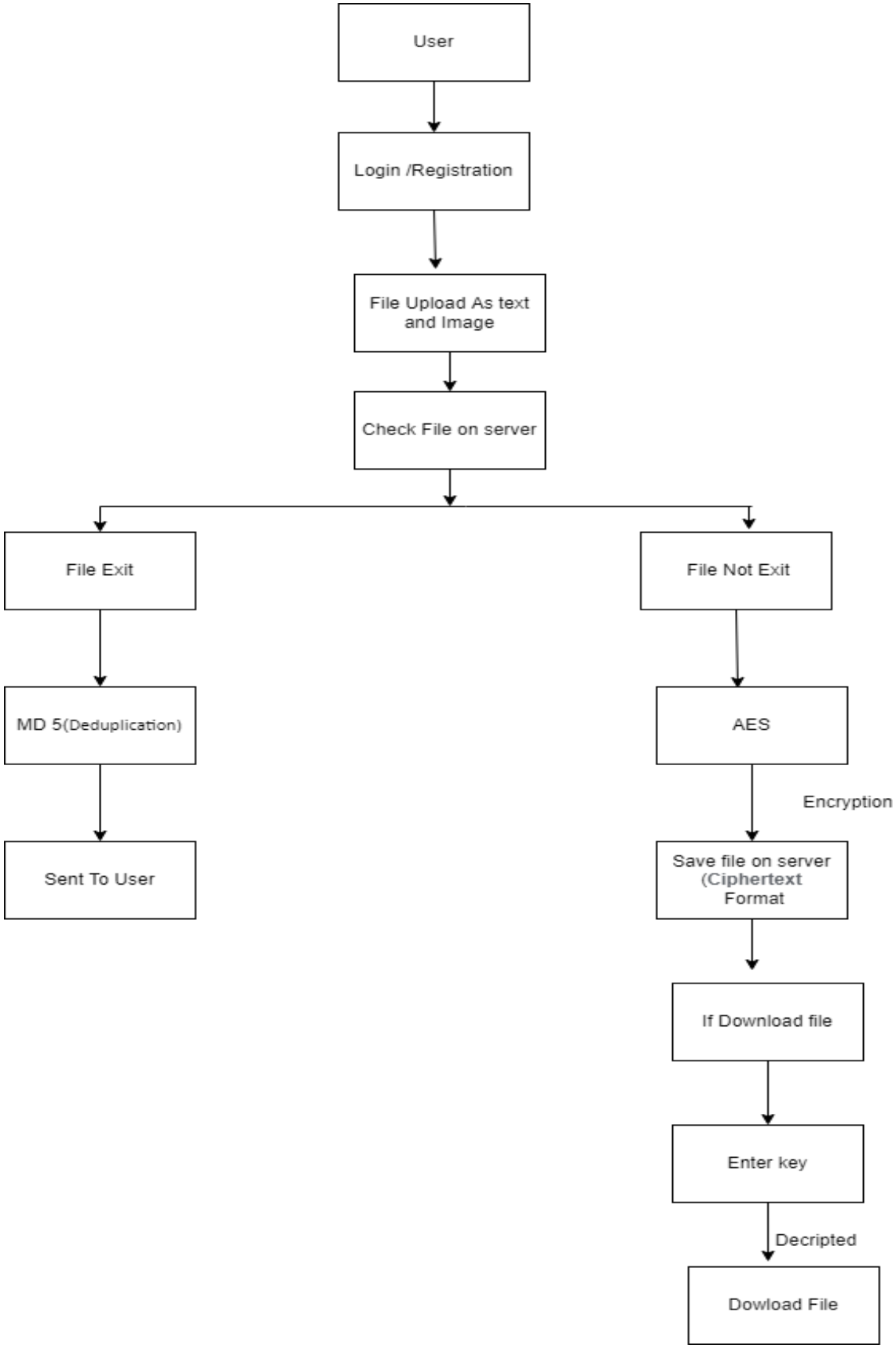

Fig. System Architecture

V. RESULTS

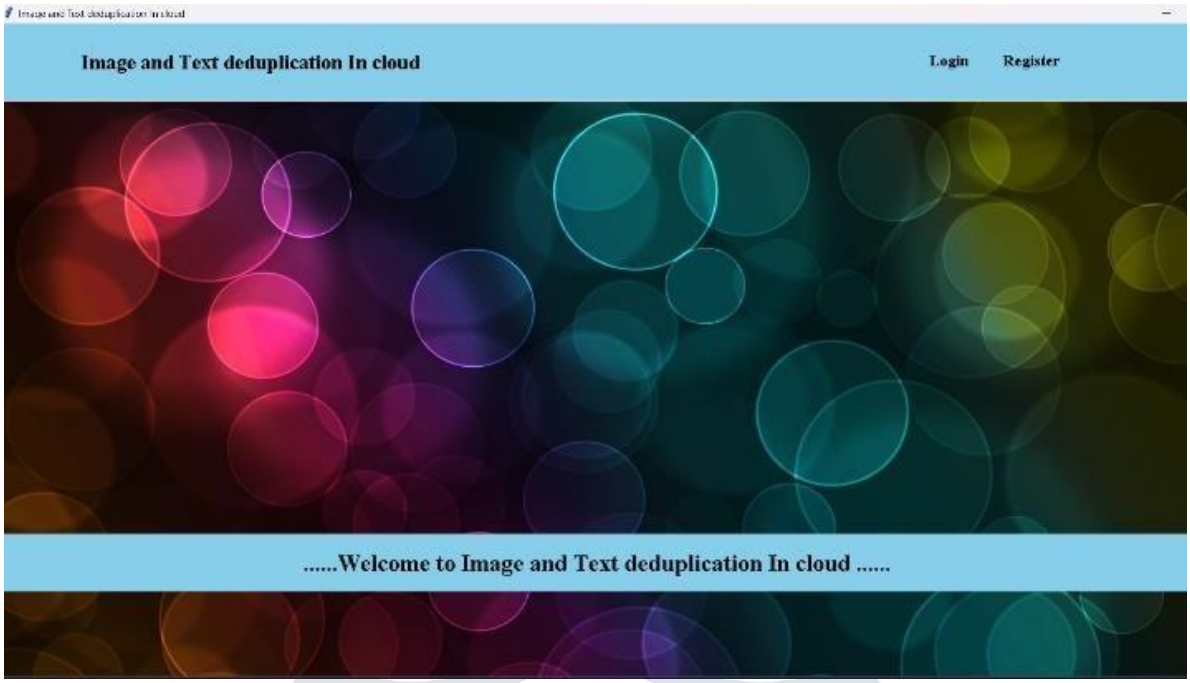

Fig. Home Page

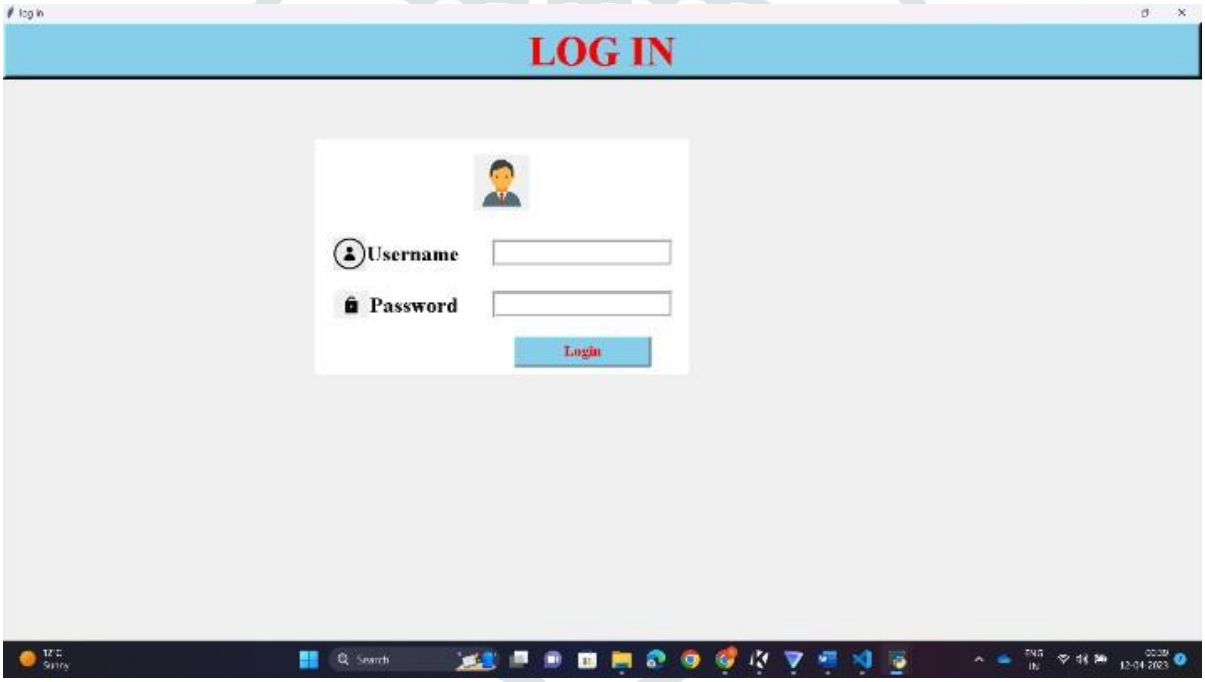

Fig. Registration and Login Page

Fig. Image Stored

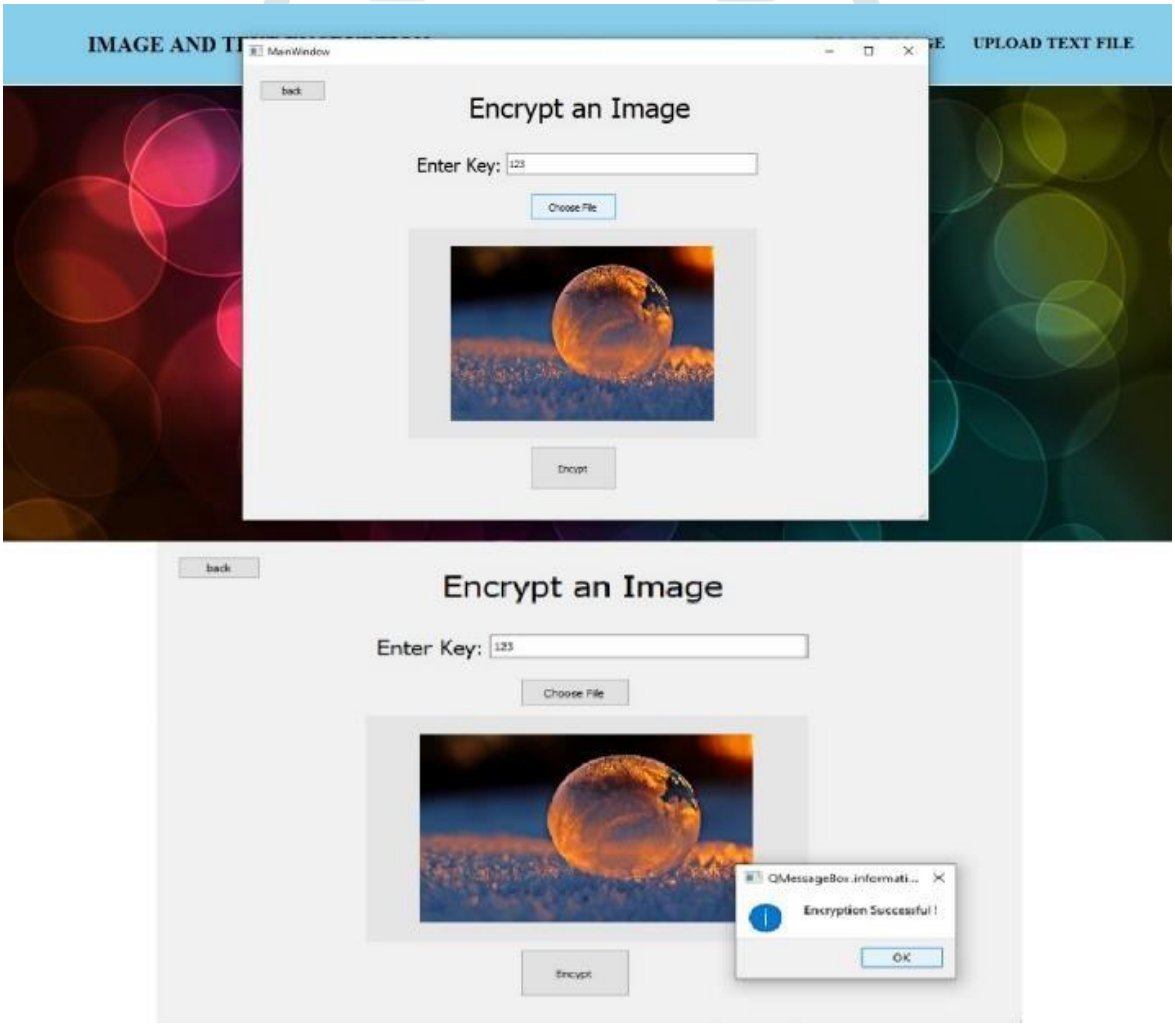

Fig. Encryption

Test Case ID	Test Case	Test Case I/P	Actual Result	Expected Result	Test case criteria(P/F)
001	Enter The Wrong username or password click on submit button	Username or password	Error comes	Error Should come	P
002	Enter the correct username and password click on submit button	Username and password	Accept	Accept	P

Test Case ID	Test Case	Test Case I/P	Actual Result	Expected Result	Test case criteria(P/F)
001	Enter the number in username, middle name, last name field	Number	Error Comes	Error Should Comes	P
001	Enter the character in username, middle name, last name field	Character	Accept	Accept	p
002	Enter the invalid email id format in email id field	Kkgmail,com	Error comes	Error Should Comes	P
002	Enter the valid email id format in email id field	kk@gmail.com	Accept	Accept	P
003	Enter the invalid digit no in phone no field	99999	Error comes	Error Should Comes	P
003	Enter the 10 digit no in phone no field	9999999999	Accept	Accept	P

Test Case ID	Test Case	Test Case I/P	Actual Result	Expected Result	Test case criteria(P/F)
001	Store Xml File	Xml file	Xml file store	Error Should come	P
002	Parse the xml file for conversion	parsing	File get parse	Accept	P
003	Attribute identification	Check individual Attribute	Identify Attributes	Accepted	P
004	Weight Analysis	Check Weight	Analyze Weight of individual Attribute	Accepted	P
005	Tree formation	Form them-Tree	Formation	Accepted	P
006	Cluster Evaluation	Check Evaluation	Should check Cluster	Accepted	P
007	Algorithm Performance	Check Evaluation	Should work Algorithm Properly	Accepted	P
008	Query Formation	Check Query Correction	Should check Query	Accepted	P

VI. CONCLUSION

Every day, a large amount of data is collected from theinternet, which needs to be protected from cybercriminals and unauthorized users through encryption. The proposed system outlines a method to prevent data duplication usingthe Message Digest 5 (MD5) algorithm. In addition, the Advanced Encryption Standard (AES) algorithm is used for data encryption and decryption. The project proposes data deduplication, which utilizes the MD5 algorithm to removeduplicate data and the AES algorithm to encrypt the data. This approach is useful for secure data deduplication, improving both storage space and data security. The project's future objectives consist of creating a cloud-basedsystem capable of managing vast quantities of data. The project relies on the Advanced Encryption Standard (AES)algorithm, known for its high level of security, for data encryption and decryption to guarantee data security and authorized access. The deduplication of data is accomplished using the MD5 algorithm, which is alsocalled Message Digest 5. In comparison to other algorithms, both AES and MD5 algorithms are recognizedas among the most secure options available.

ACKNOWLEDGMENT

We are pleased to present the initial project report on “Efficient and Secure Cloud Storage: Password-Protected Encryption. Keys for Deduplication Systems”. We would like to express my gratitude to our internal guide, Prof. S. S. Bhosale, for providing us with all the necessary assistance and guidance. Their helpful suggestions were very much appreciated. We would also like to thank Dr. S. K. Jagtap, Head of Electronics and Telecommunications Engineering Department, SKNCOE, for her invaluable support and suggestions. We are grateful to our Principal, Dr. A. V. Deshpande, for his constant support and motivation, which greatly contributed to the success of this project.

REFERENCES

[1] S. Uthayashangar, J. Abhinaya, V. Harshini, R. Jayavardhani “Image and Text Encrypted Data with Authorized Deduplication in Cloud”.

[2] Gulsayyar Ali, Dr. Mian Ilyas Ahmad, Arslan Rafi “Secure Block-level Data Deduplication approach for Cloud Data Centers”.

[3] Chippy Jacob, Rekha V. R “Secured and Reliable File Sharing System with Deduplication USING Erasure Correction Code”.

[4] W, K. Ng. Y. Wen, and H. Zhu, “Private data deduplication protocols in cloud storage” in Proceeding of the 27th Annual ACM Symposiums on Applied Computing; ACM, 20 12, pp. 441-446.

[5] R. Oi Pietro and A. Sorniotti. “Boosting efficiency and security in proof of ownership for deduplication” in Proceedings of the 7th ACM Symposium on Intonation. Computer and Communications Security. ACM, 2012, pp. 81-82.

[6] W, K. Ng. Y. Wen, and H. Zhu, “Private data deduplication protocols in cloud storage” in Proceeding of the 27th Annual ACM Symposiums on Applied Computing; ACM, 20 12, pp. 441-446.

[7] Xu, E.-C. Chang, and l. Zhou, “Weak leakage-resilient client-side deduplication of encrypted data in cloud storage” in Proceedings of the Sd! ACMSIGSAC symposium on Information, computer and communications security. ACM. 20 J3, pp.

[8] M. Li. C. Qin, and P. P, C. Lee, “Cdstore; toward reliable, secure and cost-efficient cloud storage via convergent dispersal” in usenix Technical Conference, 2015, pp. 45-53.

[9] J. Blasco, R. Di Pietro, A. Orfila, and A. Sorniotti. “A tunable proof of ownership scheme for deduplication using bloom filters,” in Comm1ications and Network Security (eNS). 2014 IEEE Conference on IEEE.