www.ijprems.com
editor@ijprems.com

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

Vol. 03, Issue 05, May 2023, pp : 1308-1312

e-ISSN : 2583-1062

Impact Factor : 5.725

TEXT AND IMAGE DEDUPLICATION IN CLOUD

Shubham Gaikwad[1], Sanket Gandhi[2], Bhupesh Bharambe[3], Pradip Rathod[4], Prof. Anoop kumar Kushwaha[5]

[1,2,3,4]CSE Department, Pune Padmbhooshan Vasantdada patil institute of technology, Pune

[5]Assistant Professor, CSE Department, Pune Padmbhooshan Vasantdada patil institute of technology, Pune India

ABSTRACT

The great development of cloud computing in recent years, the explosive increasing of image and text data, the mass of information storage, and the application demands for high availability of data, network backup is facing an unprecedented challenge. Image and text deduplication technology is proposed to reduce the storage space and costs. To protect the confidentiality of the image and text , the notion of convergent encryption has been proposed. In the deduplication system, the image and text will be encrypted/decrypted with a convergent encryption key which is derived by computing the hash value of the image content . It means that identical image copies will generate the same ciphertext, which used to check the duplicate image copy. Security analysis makes sure that this system is secure.

INDEX TERMS- Introduction, overview Objective, Algorithm, System Architecture, project demonstration, output, Future scope, Conclusion, Literature survey, references.

1. INTRODUCTION

Ith s social media grows in popularity and use, people are posting, sharing, and sending data in record numbers. The majority of software apps, social media sites, and businesses utilize cloud services to store their massive amounts of data. Files with the same content might be uploaded by the same or different users, causing the system to store the same files again and over, wasting the relatively costly storage space purchased from cloud service providers. Existing cloud storage companies de-duplicate data to minimize wasting space, which benefits both themselves and their consumers. Deduplication may save backup storage requirements by up to 9095 percent [11] and regular file system storage requirements by up to 68 percent. Encrypting the same files with different keys entered by users results in the generation of different cypher messages, even though the underlying plain text is the same. 1 As a result, classical encryption fails in data de-duplication on encrypted files. However, encryption is expected to protect the security and secrecy of data. Previous de-duplication technologies, however, cannot guarantee the data's robustness. Furthermore, many de1duplication technologies require the data owner to all be brought online 1 in order to exchange a convergence key, therefore decryption cannot be performed just at time it is requested. Previous systems did not address storage server assaults and data retrieval in such attacks. In this research, 1 we propose a de-duplication method which is based on an erasure correction technique that splits the file into shards and distributes it over several cloud storage providers' servers. Even if only one of the servers is attacked by an intruder, the system can re-generate the original files using the remaining of repaired shards. Like a outcome, the system can guarantee the encrypted file's dependability and robustness.

2. METHODOLOGY

This research paper explains how we four students completed this task within 6-7 moths of period.me and my team members gave their 100% to complete this project. We use cloud technology and encryption technics to complete this project. And then we gave it for testing for every member of group. We have completed this project with pursuing our final year engineering course.

Text and Image data It has to keep secure in a cloud server. Digital images have to be protected over the communication, however generally personal identification details like copies of pan card, Passport, ATM, etc., to store on one’s own pc. So, we are protecting the text file and image data for avoiding the duplication in our proposed system. emergence of cloud storage service, managing business/personal data via a cloud storage provider such as Dropbox, OneDrive and Google Drive has become a common option.

To construct a system with all mentioned merits, a trivial solution is to simply combine all existing techniques. For instance, the leaders of the corporation release a document of regulation, all employees will download, learn and then store it under their own accounts. If cloud service providers choose one of these schemes as the core technique of the cloud system, additional independent modules must be deployed simultaneously in order to obtain functionalities unrealized by the scheme. Then besides the significant increase on the storage, computation and communication cost, extra adjustment is needed for letting all modules collaborate as a whole. All interfaces and parameters should be correctly docked and all parameter

3. OBJECTIVE

Encryption leverages advanced algorithms to encode the data making it meaningless to any user who does not have the key Authorized users leverage the key to decode the data transforming the concealed information back into a readable format. Then besides the significant increase on the storage, computation and communication cost, extra adjustment is needed for letting all modules collaborate as a whole. All interfaces and parameters should be correctly docked and all parameters should be well adjusted.

4. LITERATURE SURVEY

[1] In this Paper "Secondary Encrypted Secure Transmission in Cognitive Radio Net- works" The Authors have Proposed Dawei Wang; Pinyi Ren; Qian Xu; Qinghe Du In order to secure the primary privacy information and provide quality- of-service provisioning for the secondary system, we propose a secondary encryption secure transmission scheme. In the proposed scheme, the primary system utilizes the secure secondary messages to encrypt the primary confidential messages and the secondary system can acquire some spectrum opportunities. Specifically, when the primary system is secure, the primary information can be directly transmitted; when the primary system is insecure while the secondary messages can be securely trans- mitted, the primary system utilizes the secure secondary messages to encrypt the primary information; otherwise, the spectrum will be utilized for secondary trans- mission. For the proposed scheme, we investigate the performances of the primary ergodic secrecy rate and the average secondary throughput. Numerical results have demonstrated that the secondary encryption secure transmission scheme can secure the primary privacy messages and improve the secondary transmission throughput.

[2] In this paper "3D-Playfair Encrypted Message Verification Technology The authors Wen-Chung Kuo; Wan-Hsuan Kao; Chun-Cheng Wang; Yu-Chih Huang have proposed that ,In the world of information development, the transmission of information is much more convenient. However, the transmission process always faces the risk of being attacked, stolen and tampered. For this reason, some scholars proposed to protect important information in the form of passwords .This paper uses 3D-Playfair encryption for encryption. However, simple 3D-playfair encryption cannot guarantee the integrity of data during transmission, so the author proposes Combined with MD5 to ensure the integrity of the data, but there are doubts about the credibility of the data source, so this paper uses XOR calculation methods to further verify the credibility of the data. When a man-in-the-middle attack is encountered, the attacker intercepts the packet And tampering with the data content can still accurately determine whether the source of the data is the original sender. This method guarantees the integrity of the data while improving the credibility of the data.

[3] In this paper "3D-Playfair Encrypted Message Verification Technology "

The authors Wen-Chung Kuo; Wan-Hsuan Kao; Chun-Cheng Wang; Yu-Chih Huang have proposed that ,In the world of information development, the transmission of information is much more convenient. However, the transmission process always faces the risk of being attacked, stolen and tampered, which leads to the doubt that the data source is incorrect. For this reason, some scholars proposed to protect important information in the form of passwords. Alok et al. Proposed 3D-Playfair Cipher with Message Integrity using MD5. This paper uses 3D-Playfair encryption for en cryption. However, simple 3D-playfair encryption cannot guarantee the integrity of data during transmission, so the author proposes Combined with MD5 to ensure the integrity of the data, but there are doubts about the credibility of the data source, so this paper uses XOR calculation methods to further verify the credibility of the data. When a man-in-the-middle attack is encountered, the attacker intercepts the packet And tampering with the data content can still accurately determine whether the source of the data is the original sender. This method guarantees the integrity of the data while improving the credibility of the data.

[4] In this paper "Dual Protection on Message Transmission based on Chinese Remainder Theorem and Rivest Cipher 4"The Authors Kevin Ronaldo Cahyono; Christy Atika Sari; De Rosal Ignatius Moses Setiadi; Eko Hari Rachmawanto Have Proposed that This research proposes a combination of dual protection on text messages transmission using Chinese Remainder Theorem (CRT) steganography and Rivest Cipher 4 (RC4) encrypting method. This combination aims to optimize the performance of encryption and message insertion into an image. Security This message is done by encrypting text messages using RC4 first, then the results are embedded in the grayscale type container image with the CRT method. The evaluation standards that will be used in this research are Mean Square Error (MSE), Peak Signal to Noise Ratio (PSNR), Structural Similarity Index Metric (SSIM), and Character Error Rate (CER). MSE, PSNR and SSIM are used as a measure of the quality of stego images. To determine the performance of the proposed method, message insertion is carried out in three types of sizes, namely maximum payload, half payload and one quarter payload. While the CER is used to find out the results of decryption of text messages. The resulting CER value is 0, this indicates the message was extracted and decrypted perfectly.

5. PROPOSED SYSTEM

To The working of the proposed is based on the fact that the texts present in images have some unique features we use Following Module In proposed System .

- **Deduplication**: For the Deduplication we use MD5 Algorithm. If deduplication Occur in file then we sent to user again and if file contain is not deduplication then store file.
- **Encryption**: File contain is unique That time AES algorithm working and store file in encrypted format
- **Decryption**: If user want or access file or download file in original format that time AES algorithm download file in decrypted format.

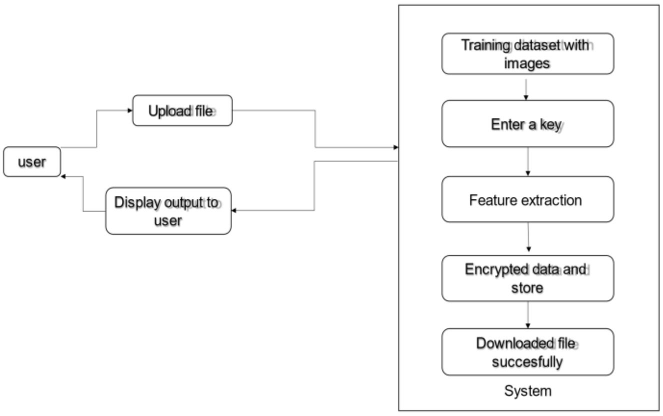
6. ALGORITHM

AES- The AES algorithm (also known as the Rijndael algorithm) is a symmetrical block cipher algorithm that takes plain text in blocks of 128 bits and converts them to ciphertext using keys of 128, 192, and 256 bits. Since the AES algorithm is considered secure, it is in the worldwide standard. The Advanced Encryption Standard (AES) is a symmetric block cipher chosen by the U.S. government to protect classified information. AES is implemented in software and hardware throughout the world to encrypt sensitive data. It is essential for government computer security, cybersecurity and electronic data protection.

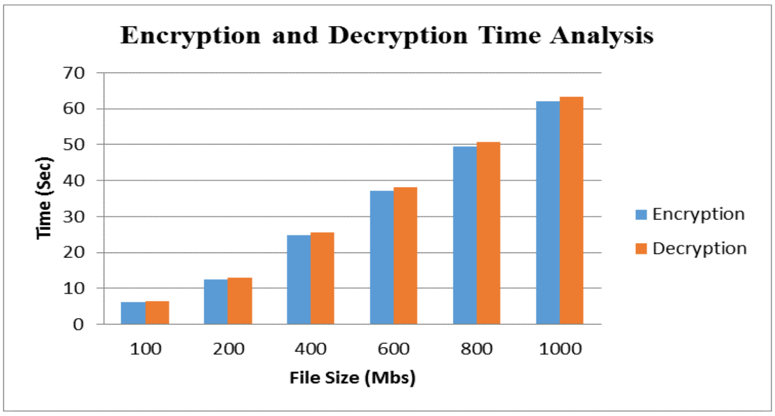
MD5- The MD5 message-digest algorithm is a cryptographically broken but still widely used hash function producing a 128-

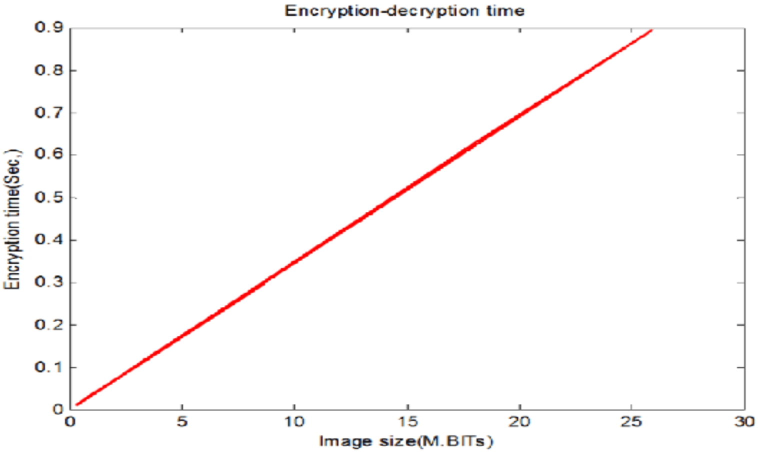
bit hash value. Although MD5 was initially designed to be used as a cryptographic hash function, it has been found to suffer from extensive vulnerabilities.

7. SYSTEM ARCHITECTURE



8. IMAGES / OUTPUT





9. UPLPOAD AN IMAGE

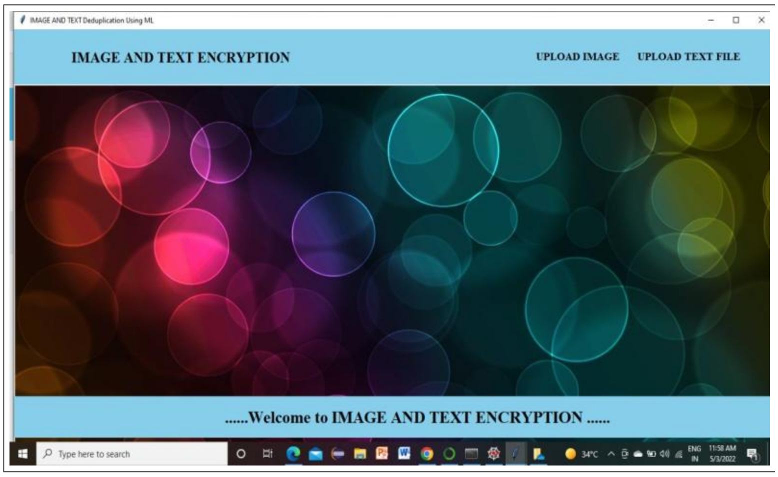
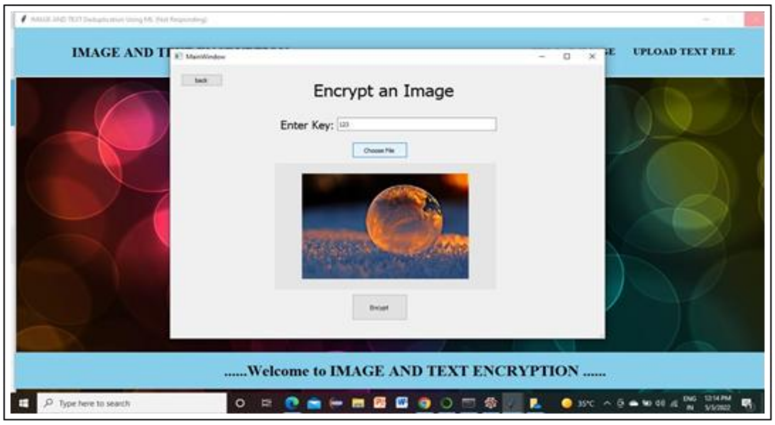


IMAGE AND TEXT Deduplication Using ML
IMAGE AND TEXT ENCRYPTION
UPLOAD IMAGE
UPLOAD TEXT FILE
......Welcome to IMAGE AND TEXT ENCRYPTION ......
Type here to search

ENCRYPT AN IMAGE WITH THE KEY



Enter Key: 123

ENCRYPTION SUCCESFUL

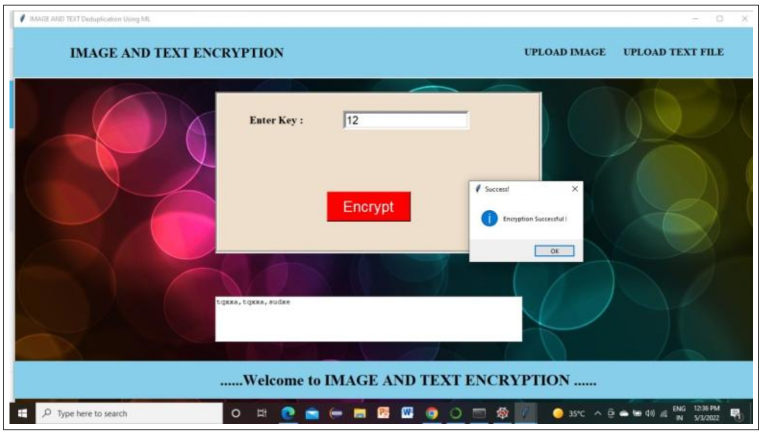


IMAGE AND TEXT ENCRYPTION
UPLOAD IMAGE
UPLOAD TEXT FILE
Encrypt
Success!
OK
......Welcome to IMAGE AND TEXT ENCRYPTION ......
Type here to search

Enter Key : 12

Encryption Successful !

tqxxa,tqxxa,sudxe

10. FUTURE SCOPE

1.Storage optimization

Deduplication helps eliminate redundant data, reducing the storage space required for storing images and text documents in the cloud. This optimization can lead to cost savings, especially for organizations dealing with large volumes of data.

2. Security and Privacy:

Deduplication can contribute to enhancing security and privacy in the cloud. By eliminating duplicate files, it reduces the surface area for potential attacks and helps prevent unauthorized access to sensitive information. Additionally, deduplication techniques can be used to detect and remove personally identifiable information (PII) from duplicated content, ensuring compliance with privacy regulations.

3. Machine Learning and AI:

Image and text deduplication can facilitate training and optimization of machine learning models. Removing duplicate or similar data points ensures that models are trained on diverse and unique samples, leading to more robust and accurate predictions.

11. CONCLUSION

In this paper we discussed that to avoid the duplication using the Encryption And decryption method. And for the text uploading we are using three algorithm., For the uploading in the cloud system we are using the Structural Similarity AES Algorithmand the main purpose of the similarity index is to check the image quality such as luminance, contrast and structure, then it measures the similarity of two image. To store large amount of data with efficiency, to avoid the duplicate text and image weare using the encryption method .

12. REFERENCES

[1] S. Halevi. D. Hornik. B. Pinkos. and A. Shulman-Peleg. "Proofs of ownership in remote storage systems,"" in Proceedings of the 18th ACM SIGSAC Conference on Computer and Communications Security. ACM, 20 I I, pp. 491-500

[2] Gonzalez-Manzano and A. Orfila. "An efficient confidentiality preserving proof of ownership for deduplication," Journal of Network and Computer Applications. vol. 50, pp. 49-59, 2015.

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

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

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