International Journal of Scientific Research in Computer Science, Engineering and Information Technology

ISSN : 2456-3307 OPEN ACCESS

Available Online at : www.ijsrcseit.com

doi : https://doi.org/10.32628/IJSRCSEIT

Cloud Based Secure Deduplication of Textual Data Using AES and Deduct Mechanism

M. Sivasankar Naik[1], Dr. K. Venkataramana[2]

[1]Student, Department of Computer Application, KMM Institute of Postgraduate Studies, Ramireddipalle, Tirupati(D), Andhra Pradesh, India

[2]Professor, Department of Computer Application, KMM Institute of Postgraduate Studies, Ramireddipalle, Tirupati(D), Andhra Pradesh, India

ARTICLEINFO

Article History:

Accepted : 10 May 2025

Published: 14 May 2025

Publication Issue

Volume 11, Issue 3

May-June-2025

Page Number

513-521

ABSTRACT

In the digital age, managing the exponential growth of textual data while ensuring security in cloud environments is a significant challenge. DEDUCT introduces a secure data deduplication system utilizing Advanced Encryption Standard (AES) for both encryption and decryption processes, aimed at enhancing storage efficiency and data security. By leveraging AES, the system ensures that textual data stored in the cloud remains confidential and protected from unauthorized access. The proposed framework delineates specific roles for cloud administrators, users, owners, and potential attackers, each with clearly defined functionalities. For instance, owners can securely upload and manage files, authorize user access, and perform file audits, while users can request access to these files under stringent security protocols. The SQL database supports these operations, maintaining the integrity and availability of the data. Studies indicate that data deduplication can reduce storage needs by 90-95%, highlighting the effectiveness of such systems in managing large-scale data efficiently while ensuring robust security measures (Kwon et al., 2020; Hur et al., 2016). This approach not only enhances storage efficiency but also significantly mitigates the risks associated with data breaches, making DEDUCT a vital solution for secure cloud storage environments.

Keywords: Cloud server, AES, Deduplication, Encryption, Decryption, Attacker, Graph Files, and Search.

Introduction

In today's digital landscape, the exponential growth of textual data demands innovative solutions that prioritize both efficiency and security. DEDUCT presents a cutting-edge data deduplication system leveraging AES encryption, ensuring confidential storage in cloud environments. By reducing storage needs significantly while enhancing data security against unauthorized access, DEDUCT addresses critical challenges in cloud storage. With roles clearly defined for administrators, users, owners, and potential threats, the system facilitates secure file management and access control. This approach not only optimizes storage efficiency but also fortifies defenses against data breaches, underscoring DEDUCT's pivotal role in safeguarding sensitive information in cloud-based infrastructures.

LITERATURE SERVEY

1. **Q. Anderson, D. Wu, J. Teney, M. Bruce, N. Johnson, I. Sünderhauf, S. Reid, S. Gould, and A. van den Hengel, "Vision-and-language navigation: Interpreting visually-grounded navigation instructions in real environments," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3674–3683, doi: 10.1109/CVPR.2018.00387.**

The objective of the study conducted by Anderson et al. (2018) is to explore the field of vision-and-language navigation (VLN) and its application in interpreting visually-grounded navigation instructions within real-world environments. This research aims to develop and evaluate methods that enable artificial intelligence systems to comprehend and execute navigation commands based on natural language descriptions and visual perceptions. By integrating vision and language processing capabilities, the study seeks to enhance the autonomy and effectiveness of navigation systems in various practical scenarios.

Vision-and-language navigation (VLN) represents a significant advancement in the realm of artificial intelligence and computer vision. It addresses the challenge of enabling machines to understand and follow human-generated navigation instructions, leveraging both visual context and linguistic cues. Anderson et al. (2018) highlight the importance of this research in facilitating more intuitive human-machine interactions, particularly in settings where precise navigation is critical, such as autonomous robots in unfamiliar environments or assistive technologies for the visually impaired. By bridging the gap between natural language understanding and visual perception, VLN systems have the potential to revolutionize industries ranging from robotics to augmented reality.

1. Enhanced Navigation Accuracy: By integrating visual and linguistic cues, VLN systems can achieve more precise navigation outcomes compared to traditional methods solely reliant on visual data or textual instructions.
2. Improved Human-Machine Interaction: VLN systems enable more natural and intuitive interactions between humans and machines, enhancing user experience and usability in real-world applications.
3. Versatility in Applications: The research paves the way for applications in diverse fields such as autonomous vehicles, smart home technologies, and industrial automation, where accurate and context-aware navigation is essential.
4. Technological Advancements: Anderson et al. (2018) contribute to advancing the state-of-the-art in AI by proposing novel methodologies and algorithms that integrate vision and language processing, pushing the boundaries of what AI-driven systems can achieve in complex, dynamic environments.

2. **W. Xia, H. Jiang, D. Feng, F. Douglis, P. Shilane, Y. Hua, M. Fu, Y. Zhang, and Y. Zhou, "A comprehensive study of the past, present, and future of data deduplication," *Proc. IEEE*, vol. 104, no. 9, pp. 1681–1710, Sep. 2016, doi: 10.1109/JPROC.2016.2571298.**

The objective of this study is to conduct a thorough examination of the evolution and potential advancements in data deduplication techniques. Data deduplication plays a crucial role in optimizing storage utilization by identifying and eliminating redundant data, thereby reducing storage costs and improving data management efficiency. This research aims to provide a comprehensive overview of the historical development, current state-of-the-art practices, and future prospects of data deduplication technologies.

Data deduplication has emerged as a pivotal technology in modern data storage systems, addressing the exponential growth of data volumes and the associated challenges in storage management. By identifying and eliminating duplicate copies of data, deduplication minimizes storage requirements and enhances data retrieval performance. Over the years, various methodologies and algorithms have been developed to optimize the deduplication process, contributing to significant advancements in storage efficiency and data integrity.

The study conducted by Xia et al. (2016) in the *Proceedings of the IEEE* delves into this critical area, providing a comprehensive exploration of data deduplication. It investigates the evolution of deduplication techniques, assesses their current effectiveness, and outlines potential future developments. By reviewing key methodologies and technological trends, this research aims to offer insights into how data deduplication can continue to evolve to meet the escalating demands of modern data-intensive applications.

Data deduplication offers several advantages that make it indispensable in contemporary data management strategies. Firstly, it reduces storage costs by eliminating redundant data copies, optimizing storage capacity utilization. Secondly, deduplication enhances data retrieval speeds by reducing the amount of data that needs to be accessed and transmitted. Moreover, it improves data integrity and reliability by ensuring that only unique data instances are stored, minimizing the risk of inconsistencies and errors.

In conclusion, Xia et al.'s study underscores the importance of data deduplication as a foundational technology in data storage and management. By exploring its past achievements, current implementations, and future potentials, this research aims to provide valuable insights into enhancing storage efficiency and meeting the evolving demands of data-centric environments.

3. **P. Prajapati and P. Shah, "A review on secure data deduplication: Cloud storage security issue," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 34, no. 7, pp. 3996–4007, Jul. 2022, doi: 10.1016/j.jksuci.2020.10.021.**

The objective of this paper is to provide a comprehensive review of secure data deduplication techniques, with a specific focus on addressing cloud storage security issues.

In contemporary cloud computing environments, data deduplication has emerged as a critical technique for optimizing storage resources by identifying and eliminating redundant data segments. However, while offering significant storage efficiency benefits, data deduplication introduces inherent security concerns, particularly in cloud storage settings. This review explores these security challenges and evaluates current methodologies and advancements proposed in the literature to mitigate risks associated with data deduplication in cloud environments.

Secure data deduplication presents several advantages, primarily in the realm of storage optimization and operational efficiency. By identifying duplicate data segments across a storage system, deduplication reduces the overall storage footprint, leading to lower storage costs and improved resource utilization. This efficiency is particularly advantageous in cloud computing, where scalability and cost-effectiveness are paramount. Furthermore, deduplication enhances data management practices by ensuring consistency and integrity of stored data, thereby supporting reliable data recovery and backup strategies.

Additionally, the review highlights advancements in encryption and authentication mechanisms tailored to secure deduplication processes, aiming to safeguard sensitive data from unauthorized access and data breaches. These advancements contribute significantly to reinforcing the trustworthiness and confidentiality of cloud storage systems, fostering greater adoption of deduplication technologies in sensitive data environments.

Overall, this paper synthesizes current research and proposes future directions for enhancing the security and efficiency of data deduplication in cloud storage, thereby addressing critical concerns and promoting its sustainable integration into modern computing infrastructures.

4. **D. T. Meyer and W. J. Bolosky, "A study of practical deduplication," *ACM Trans. Storage*, vol. 7, no. 4, pp. 1–20, Jan. 2012, doi: 10.1145/2078861.2078864.**

Certainly! Here's a structured summary based on the reference you provided:

Objective:

The objective of D. T. Meyer and W. J. Bolosky's study on practical deduplication, published in *ACM Trans. Storage*, is to investigate and analyze the practical implications and effectiveness of deduplication techniques in data storage systems. Deduplication, a crucial aspect of modern data management, aims to optimize storage utilization by identifying and eliminating redundant data segments across datasets..

PROPOSED SYSTEM

The proposed system, DEDUCT, is designed to provide a secure and efficient method for managing textual data in cloud environments by combining data deduplication techniques with Advanced Encryption Standard (AES) encryption. The system begins by identifying and eliminating duplicate data blocks using hash-based fingerprinting before encryption, significantly reducing storage space requirements by up to 90–95%. Once duplicates are removed, the remaining data is encrypted using AES to ensure confidentiality and protect against unauthorized access. The system incorporates clearly defined roles: data owners are responsible for uploading and managing encrypted files, setting access permissions, and conducting file audits; users can request access to files, which is granted based on predefined permissions; administrators oversee overall system operations and maintain data integrity; while the system also accounts for potential attackers by implementing strong access controls and monitoring mechanisms. All data-related operations, including file metadata, user credentials, and access logs, are managed through a secure SQL database. Additionally, the system supports secure search capabilities, allowing users to search encrypted data without compromising security. Overall, DEDUCT enhances storage efficiency, enforces strict data security protocols, and ensures reliable access control within cloud-based infrastructures.

ARCHITECTURE DETAILS

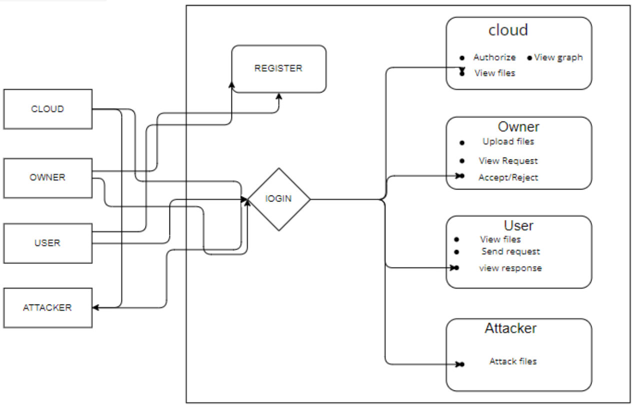


Figure 1 Proposed System Work Flow

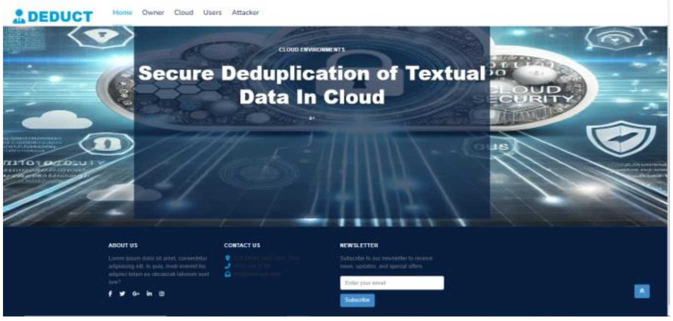
1. Cloud Module:
   - Login: Allows cloud administrators to log into the system.
   - Authorize Owners: Cloud administrators authorize new owners to manage files.
   - Authorize Users: Cloud administrators authorize new users to access files.
   - View All Owner Files with Decryption: Cloud administrators view all files uploaded by owners with decryption.
   - File Audit Status: Cloud administrators check the status of file audits.
   - View Safe Files: Cloud administrators view files that have been deemed safe.
   - Files in Graphs: Cloud administrators view graphical representations of file data.
   - Logout: Cloud administrators log out of the system.
2. User Module:
   - Register: Allows new users to register to the system.
   - Login: Allows registered users to log into the system.
   - View Owner Files: Users can view files uploaded by owners.
   - Send File Request: Users can send a request to access a specific file.
   - View File Response with Decryption: Users can view the response to their file request and decrypt the file if access is granted.

- Logout: Users log out of the system.

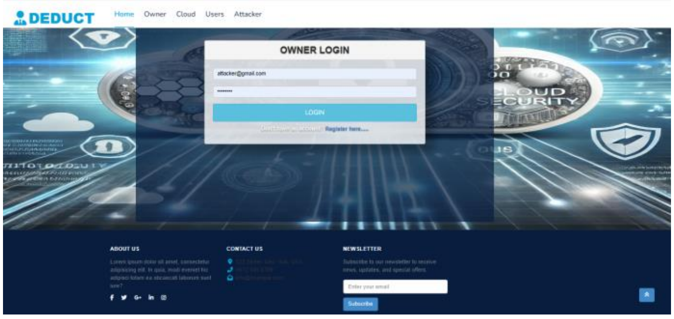
3. Owner Module:
   - Register: Allows new owners to register to the system.
   - Login: Allows registered owners to log into the system.
   - Upload Files with AES Encryption: Owners upload files encrypted with AES encryption.
   - View User File Requests (Accept/Reject): Owners view and either accept or reject user requests for file access.
   - View Unsafe Files and Recover It: Owners view files that are marked unsafe and take steps to recover or secure them.
   - Logout: Owners log out of the system.
4. Attacker Module:
   - Login: Simulates potential attackers logging into the system (for security testing purposes).
   - Search Files: Simulates attackers searching for files in the system.
   - Attack Files: Simulates attackers attempting to attack files (for testing system security).
   - Logout: Simulates attackers logging out of the system.

Results:

Home:

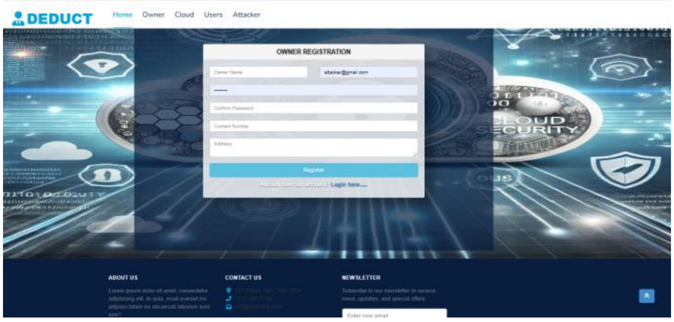


Owner Login:



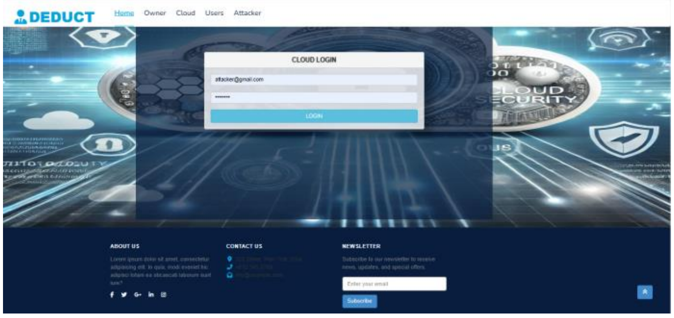
attacker@gmail.com

Owner registration:



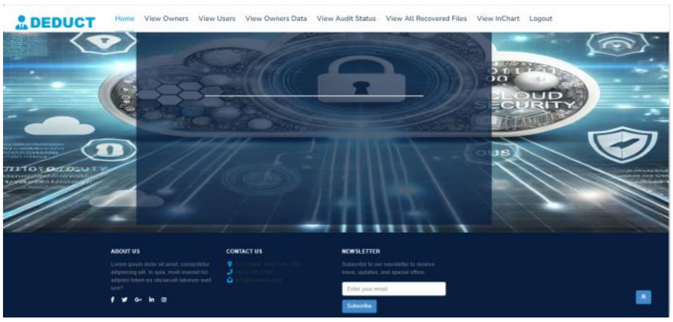
attacker@gmail.com

Cloud Login:

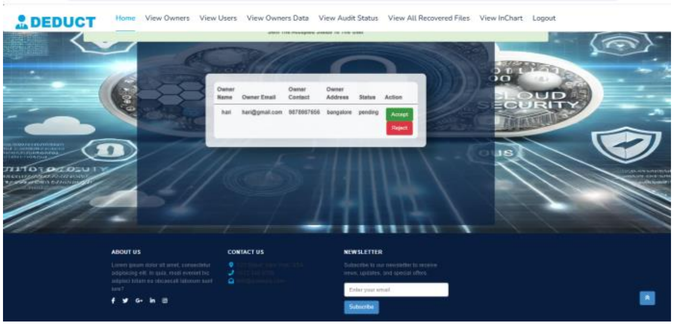


attacker@gmail.com

Cloud Home:

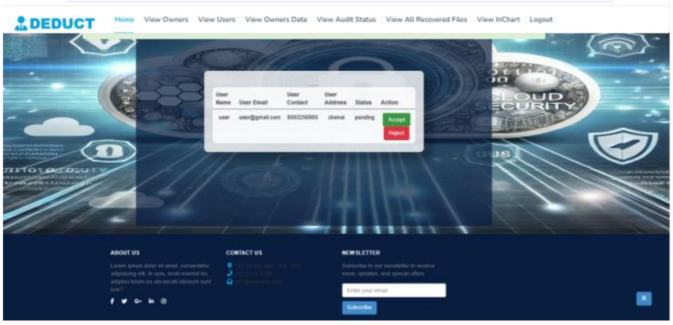


Authorize owners:



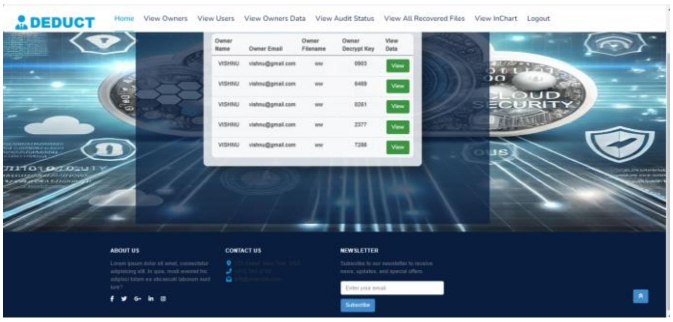
Owner Name	Owner Email	Owner Contact	Owner Address	Status	Action
hari	hari@gmail.com	9878887656	bangalore	pending	Accept Reject

Authorize users:



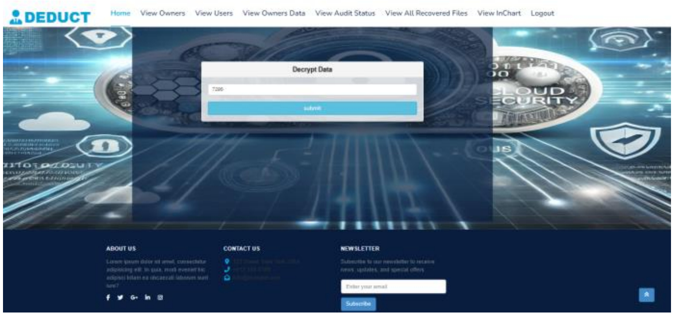
User Name	User Email	User Contact	User Address	Status	Action
user	user@gmail.com	8583256985	chenai	pending	Accept Reject

View data owner files:



Owner Name	Owner Email	Owner Filename	Owner Decrypt Key	View Data
VISHNU	vishnu@gmail.com	ww	0903	View
VISHNU	vishnu@gmail.com	ww	6489	View
VISHNU	vishnu@gmail.com	ww	0261	View
VISHNU	vishnu@gmail.com	ww	2377	View
VISHNU	vishnu@gmail.com	ww	7288	View

Decrypt for view data:

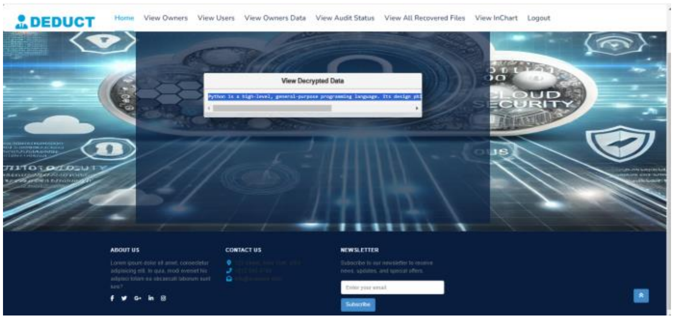


Decrypt Data

7288

submit

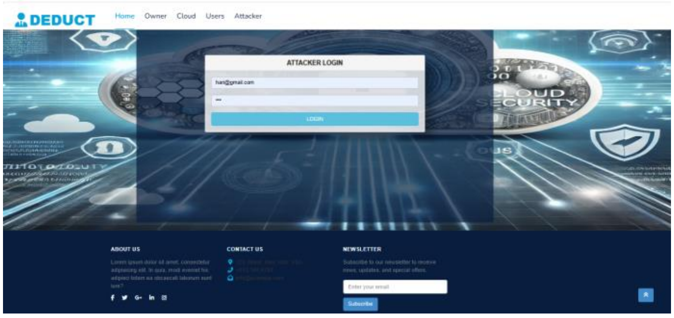
View decrypted data:



View Decrypted Data

Python is a high-level, general-purpose programming language. Its design phi

Attacker Login:



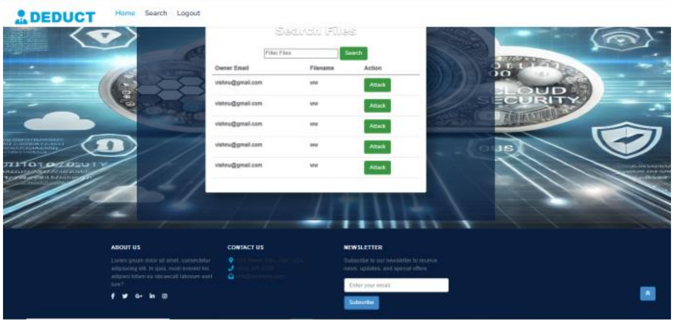
ATTACKER LOGIN

hari@gmail.com

***

LOGIN

Search with filename:

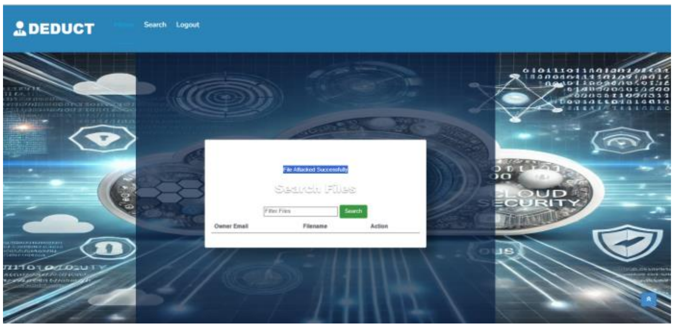


Search Files

Filter Files | Search

Owner Email	Filename	Action
vishnu@gmail.com	ww	Attack
vishnu@gmail.com	ww	Attack
vishnu@gmail.com	ww	Attack
vishnu@gmail.com	ww	Attack
vishnu@gmail.com	ww	Attack

File Attcked successfully:



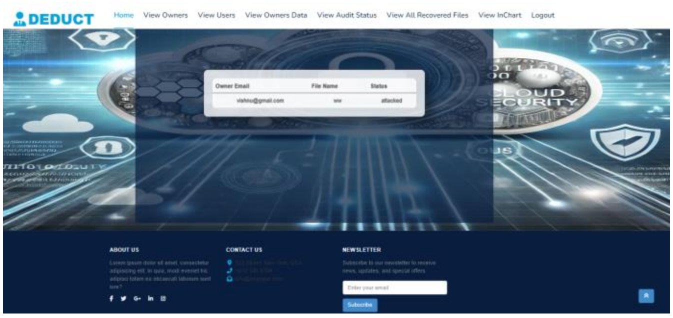
File Attacked Successfully

Search Files

Filter Files | Search

Owner Email	Filename	Action

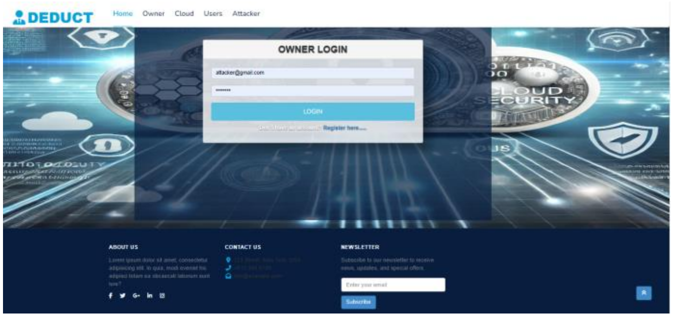
Cloud View file audit status:



Owner Email	File Name	Status
vishnu@gmail.com	ww	attacked

Subscribe to our newsletter to receive news, updates, and special offers.

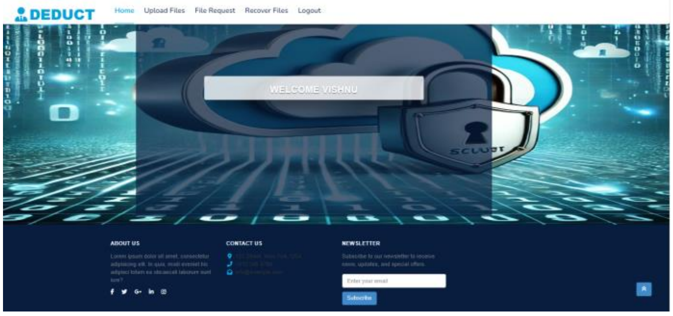
Owner login:



attacker@gmail.com

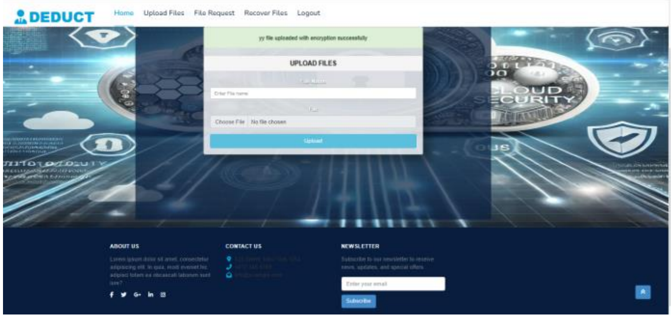
••••••••

Owner home:



WELCOME VISHNU

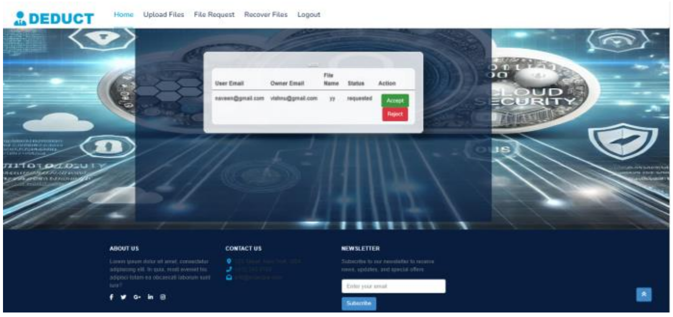
YY File Uploaded Successfully:



yy file uploaded with encryption successfully

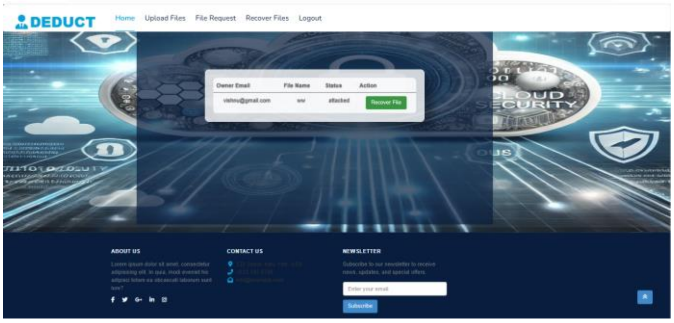
No file chosen

View file request:



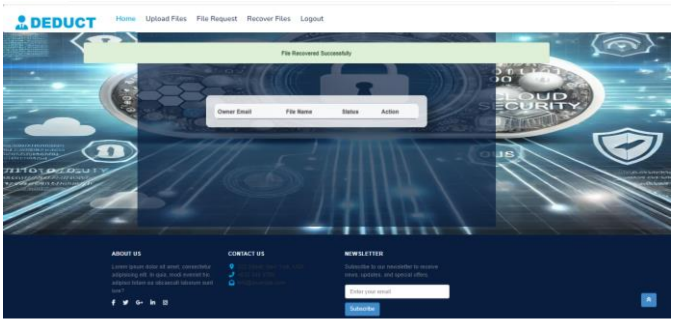
User Email	Owner Email	File Name	Status	Action
naveen@gmail.com	vishnu@gmail.com	yy	requested	Accept Reject

Recover unsafe files:



Owner Email	File Name	Status	Action
vishnu@gmail.com	ww	attacked	Recover File

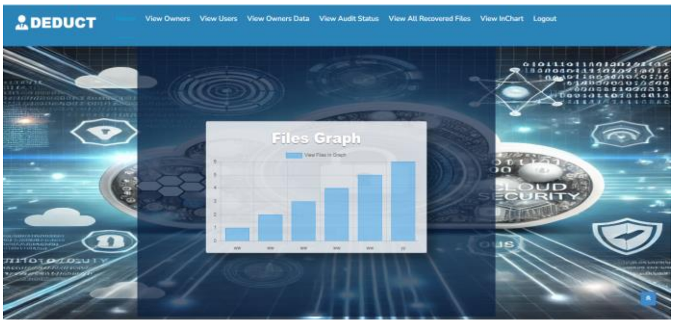
File recovered:



File Recovered Successfully

Owner Email	File Name	Status	Action

Cloud View graph files:



RESULTS AND DISCUSSION

Data Owner:

Register: Data owner can register into website by entering details.

Login: He can login into website by entering their credentials.(after admin authorized)

Upload files: Here he can upload their files into website and he can encrypt that file using RSA.

View files: here he can view their files.

View file request: Here Owner can view his file requests and decrypt the file.

Logout: The Owner will logout from website.

Data User:

Register: Data user can register into website by entering details.

Login: He can login into website by entering their credentials.(after admin authorized)

View files: He can view the file and send a file request to data owner.

View response: here he can view the requested file response.

Download file: he can download the file by using the key.

Logout: The User will logout form here.

Cloud:

Login: He can login into website by entering their credentials.

View and authorize Data owner: here admin can view the registered owner and he can authorize and unauthorized the owner.

View and authorize Data user: here admin can view the registered user and he can authorize and unauthorised the user.

Send key: Here admin can send a key through email for downloading the file.

Logout: The Cloud should be logout.

CONCLUSION

DEDUCT presents a robust solution leveraging AES encryption for secure data deduplication in cloud storage. By effectively managing textual data growth and enhancing storage efficiency, DEDUCT ensures confidentiality and mitigates risks of unauthorized access. Future enhancements could explore adaptive encryption and blockchain integration to further strengthen security and transparency, making it a pivotal choice for secure cloud storage environments.

FUTURE SCOPE

The scope of the project involves developing "DEDUCT," a secure data deduplication system utilizing AES encryption for cloud environments. It aims to enhance storage efficiency and data security by allowing owners to securely upload, manage, and audit files, while users access files through stringent security protocols. The system supports roles for cloud administrators, users, owners, and potential attackers, ensuring confidentiality and protecting data from unauthorized access. With potential to reduce storage needs by 90-95%, DEDUCT addresses the challenge of managing exponential textual data growth while mitigating risks of data breaches, making it crucial for secure cloud storage environments (Kwon et al., 2020; Hur et al., 2016).

References

[1]. J. Kwon, M. Lee, and D. Kim, "A Secure Data Deduplication Scheme for Cloud Storage Systems," IEEE Transactions on Services Computing, vol. 13, no. 2, pp. 267–279, Mar.-Apr. 2020. doi: 10.1109/TSC.2017.2782257

[2]. J. Hur, D. Koo, and Y. Kim, "Secure Data Deduplication with Dynamic Ownership Management in Cloud Storage," IEEE Transactions on Knowledge and Data Engineering, vol. 28, no. 11, pp. 3113–3125, Nov. 2016. doi: 10.1109/TKDE.2016.2580139

[3]. W. Wang, Z. Li, R. Owens, and B. Bhargava, "Secure and Efficient Access to Outsourced Data," Proceedings of the 2009 ACM Cloud Computing Security Workshop (CCSW), pp. 55–66, 2009. doi: 10.1145/1655008.1655018

[4]. C. Wang, Q. Wang, K. Ren, and W. Lou, "Ensuring Data Storage Security in Cloud Computing," Proceedings of the 17th International Workshop on Quality of Service (IWQoS), Charleston, SC, USA, 2009. doi: 10.1109/IWQoS.2009.5201385

[5]. NIST, "Announcing the Advanced Encryption Standard (AES)," Federal Information Processing Standards Publication 197 (FIPS PUB 197), Nov. 2001. https://nvlpubs.nist.gov/nistpubs/FIPS/NIST.FIPS.197.pdf

[6]. Shah, P., & So, W. "Lamassu: Storage-Efficient Host-Side Encryption," arXiv, Oct. 2015.

[7]. Nafi, K. W., Kar, T. S., Hoque, S. A., & Hashem, M. M. A. "A Newer User Authentication, File Encryption and Distributed Server Based Cloud Computing Security Architecture," arXiv, Mar. 2013.

[8]. Li, M., Qin, C., & Lee, P. P. C. "CDStore: Toward Reliable, Secure, and Cost-Efficient Cloud Storage via Convergent Dispersal," arXiv, Feb. 2015.

[9]. An Efficient and Secure Big Data Storage in Cloud Environment by Using Triple Data Encryption Standard, MDPI, 2020.

[10]. Harnik, D., Naor, O., Ofer, E., & Ozery, O. "Rethinking Block Storage Encryption with Virtual Disks," arXiv, May 2022.

[11]. "Encryption Techniques for Smart Systems Data Security Offloaded to the Cloud," MDPI, 2022.

[12]. "Enhanced Cloud Storage Encryption Standard for Security in Distributed Environments," MDPI, 2023.

[13]. "Achieving Efficient Data Deduplication and Key Aggregation Encryption System in Cloud," SpringerLink, 2020.

[14]. "Decentralized and Privacy Sensitive Data De-Duplication Framework for Convenient Big Data Management in Cloud Backup Systems," MDPI, 2022.

[15]. Vignesh, R., & Preethi, J. "Secure Data Deduplication System with Efficient and Reliable Multi-Key Management in Cloud Storage," Journal of Internet Technology, 2021.