IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

SECURE FILE ACCESS SYSTEM WITH DIPLICATION DETECTION AND MODIFICATION ALERTS

1Kamal M, 2Karthikeyan M V, 3Sivadharshan N

1B.Tech Computer Science and Business System, 2B.Tech Computer Science and Business System, 3B.Tech Computer Science and Business System

1K. Ramakrishnan College of Engineering,

2K. Ramakrishnan College of Engineering,

3K. Ramakrishnan College of Engineering

Abstract

This project introduces a real-time, AI-powered system for detecting data download duplication, modification, and authentication issues. Moving beyond traditional **hash- based integrity** checks, the system employs machine learning models to establish and monitor the behavioral fingerprint of files and their sources.

The implementation integrates a crucial **Multi Factor Authentication (MFA) layer**, requiring users to verify their identity time-sensitive security code before executing sensitive actions like overriding an AI alert or modifying system settings. Key technologies include **TensorFlow/Py Torch** for AI/ML modeling, a **PostgreSQL database** for storing file metadata and behavioral data, and a flexible alerting system for notifying users in real time. By integrating advanced analytics with a comprehensive security framework and strong user authentication, this solution aims to significantly enhance data integrity, security, and operational resilience against a wide range of cyber threats and human errors.

Index Terms – Deduplication, Data Integrity, SHA-256, Modification Detection, Multi-factor Authentication, Data Storage Enhancement.

1.Introduction

The Cybersecurity and Artificial Intelligence (AI) domain focuses on securing digital systems by using intelligent technologies that can learn, analyse , and respond to threats automatically. With the rapid growth of the internet, cloud computing, and connected devices, cyberattacks have become more frequent and advanced. Traditional security methods cannot handle the large volume of data or detect new types of attacks quickly.

AI helps overcome these challenges by identifying unusual patterns, analysing user behaviour detecting malware or phishing attempts, and predicting possible threats in real time. It improves accuracy, reduces human effort, and enables faster response during security incidents. By combining AI with cybersecurity, organizations can create stronger, adaptive, and more efficient security systems that protect networks, data, and applications from modern cyber risks. This makes the Cybersecurity + AI domain highly relevant for research and final year projects.

2.Literature Review

With the rapid growth of digital data, secure file management has become essential. Existing systems mainly focus on storage and basic security but lack effective duplicate detection and modification monitoring. Cryptographic techniques like AES are used for secure file storage, while SHA-256 hashing helps in ensuring data integrity and detecting duplicate files. However, most systems detect only exact duplicates and do not support similarity detection. Recent research introduces AI techniques such as CNN for images and MFCC for audio to identify near-duplicate content. Additionally, OTP-based authentication and real-time alerts improve access control and system security. Based on these approaches, this project integrates encryption, deduplication, modification detection, and alert mechanisms to provide a more secure and efficient file management system.

3.Research Methodology

3.1. System Design and Architecture

The proposed system is designed using a modular architecture to ensure scalability, security, and maintainability. The system consists of multiple modules such as user authentication, file upload and storage, deduplication, modification detection, access control, alert generation, and logging. A client-server model is followed, where the frontend is developed using HTML, CSS, and JavaScript, and the backend is implemented using Python (Flask) with MySQL as the database. This structured design allows each module to function independently while contributing to the overall system performance.

3.2. Data Security and Encryption Techniques

To ensure data confidentiality and integrity, strong cryptographic techniques are used. User passwords are securely stored using hashing algorithms, while uploaded files are encrypted using the Advanced Encryption Standard (AES). SHA-256 hashing is applied to generate unique hash values for files, which are used for both deduplication and integrity verification. These techniques ensure that sensitive data is protected from unauthorized access and tampering.

3.3. Deduplication and Integrity Verification

The system implements a hash-based deduplication approach to identify and eliminate duplicate files. When a file is uploaded, its SHA-256 hash is generated and compared with existing hashes in the database. If a match is found, the file is identified as a duplicate and is not stored again. Additionally, file integrity is verified by recalculating the hash during access and comparing it with the original stored hash. This helps in detecting any unauthorized modifications to the file.

3.4. Access Control, Alerts, and Monitoring

A dual-level authentication mechanism is used to enhance security, where users must pass both login authentication and OTP-based verification for file access. The system generates real-time alerts for events such as duplicate uploads, unauthorized access attempts, and file modifications. All activities are recorded in a logging module with timestamps, ensuring transparency and enabling audit trails. This approach improves system reliability and provides continuous monitoring of user actions.

4. Results and Discussion

This section evaluates the performance, storage efficiency, and security responsiveness of the proposed system. The testing was conducted under a controlled environment to measure three primary components: the efficiency of the duplication detection algorithm, the computational overhead of the security mechanisms, and the latency of the modification alert system.
Compared to traditional methods, the AI-based approach offers:
Faster processing
Higher accuracy
Reduced storage of duplicate files

5. Conclusion

The Secure File Access System with Duplication Detection and Modification Alerts provides a comprehensive and intelligent solution to overcome the growing challenges associated with file security, redundancy, and unauthorized tampering in modern digital environments. Through the integration of hashing techniques, encryption mechanisms, dual- layer authentication, and real-time alerting, the system ensures a far superior level of protection compared to traditional storage methods.

6. Future Scope

- Secure File Management & Authentication
- Duplicate & Similar File Detection
- Real-Time Monitoring & Tampering Detection

References

[1] LeCun, Y., Bengio, Y., & Hinton, G. “Deep Learning”, Nature, 2015.
[2] Krizhevsky, A. et al. “ImageNet Classification using CNN”, 2012.
[3] WHO. “Biomedical Waste Management Guidelines”.
[4] Recent research articles on AI-based waste classification.