Opaline Attachment Defence System For Proactive Detection And Sanitization of Malicious Email Files

Mrs. M.Radhika[1], G.Ashika[2], Shreya Agrawal[3], T.Sruthik[4]
[1]Assistant Professor, Dept of Information Technology
[2, 3, 4]Dept of Information Technology
[1, 2, 3, 4] R.M.D. Engineering College, Chennai, Tamil Nadu.

Abstract- *Email attachments remain a primary vector for phishing and malware attacks, with traditional signature-based defences struggling against zero-day threats. This paper introduces Opaline, a proactive defence system that leverages deep learning models like RoBERTa to analyse textual semantics and structural features in attachments such as PDFs and documents, achieving early detection before user interaction. By integrating sandbox isolation and automated sanitization—converting suspicious files into safe static previews—Opaline minimizes risks while preserving workflow usability, offering a practical advancement over resource-heavy existing solutions.*

Keywords: Email attachment security, malicious file detection, deep learning cybersecurity, RoBERTa malware classification, proactive threat sanitization, sandbox isolation, phishing prevention, zero-day attack defence, attachment preview safety, Flask-based security system.

I. INTRODUCTION

Email attachments represent a persistent and evolving threat in cybersecurity, frequently exploited by attackers to deliver phishing payloads, macros, and zero-day malware hidden within seemingly innocuous PDFs, Word documents, Excel sheets, or images. These files leverage vulnerabilities in common software viewers, enabling unauthorized system access, data theft, or ransomware deployment upon execution—issues exacerbated by the sheer volume of daily emails processed globally. Conventional defences, such as signature-based antivirus scanners and heuristic analysers, prove inadequate against novel or obfuscated threats, often resulting in delayed detection or high false negatives that leave end-users vulnerable.Motivated by the rising incidence of attachment-centric attacks documented in recent reports, this paper introduces Opaline, an innovative defence system that employs deep learning models like RoBERTa to dissect both textual semantics and structural anomalies in attachments proactively. Beyond mere classification, Opaline integrates secure sandbox isolation for suspicious files and automated sanitization—converting them into safe, static previews (e.g., images or text extracts)—thereby mitigating risks while preserving seamless email workflows. Our approach not only surpasses traditional methods in accuracy and speed but also prioritizes usability, offering a practical solution for enterprise and personal email security.

1.1 PROBLEM STATEMENT

Malicious email attachments, often disguised as legitimate PDFs, Word documents, Excel spreadsheets, or images, pose a severe cybersecurity risk by embedding scripts, macros, or exploits that compromise systems upon opening. Attackers craft highly convincing files with hidden payloads, exploiting software vulnerabilities to enable data theft, ransomware, or remote control—despite widespread use of antivirus tools and email filters. Current defences rely heavily on signature-based detection and heuristics, which fail against zero-day malware, polymorphic variants, and sophisticated phishing that evades known patterns, resulting in delayed or missed threats. Sandboxing solutions, while helpful, are resource-intensive, poorly integrated with email clients, and do not prevent user exposure during preview. This leaves a critical gap: the absence of proactive, intelligent systems that detect anomalies in attachment semantics and structure while providing safe access without execution or workflow disruption.

1.2 OBJECTIVE OF THE STUDY

The primary objective of this study is to develop Opaline, an advanced attachment defence system that proactively detects and neutralizes malicious email files using deep learning techniques, overcoming the shortcomings of conventional security measures. Specifically, we aim to analyse both textual semantics and structural features of attachments—such as PDFs, Word documents, and spreadsheets—employing models like RoBERTa within a TensorFlow framework to identify phishing attempts and embedded malware with high accuracy before user interaction. A key goal is to isolate suspicious files in a secure sandbox environment, preventing potential exploitation while automatically sanitizing them by converting dynamic content into safe, static formats like images or plain text extracts for

risk-free preview. Ultimately, Opaline seeks to enhance real-time email security, reduce false negatives against zero-day threats, and maintain usability for both enterprise and individual users, fostering safer digital communication in an era of escalating cyber risks.

II. LITERATURE REVIEW

Several studies have explored job recommendation, skill gap identification, and career guidance using machine learning and data-driven techniques. The following works are closely related to the proposed system.

In paper [1],Alsuhibany et al.compares BERT and RoBERTa for email spam classification; RoBERTa shows superior recall via semantic analysis of email content.

In paper [2], Mahfoud et al. proposes ensemble deep learning for malicious emails using headers, body, attachments; achieves 0.993 AUC on diverse datasets.

In paper [3], Vinayakumaret al.uses RoBERTa-large with metadata projection for malicious URL detection in phishing; reports 98% accuracy.

In paper[4], Smith leverages dynamic binary instrumentation and DL for zero-day malware; Pulse/Alpha frameworks hit 96-100% detection rates.

In paper [5],OPSWAT Blog advocates proactive email security via attachment sanitization over reactive scanning; blocks macros pre-delivery effectively.

III. SYSTEM ARCHITECTURE

Opaline's architecture features a Flask web interface for email attachment uploads, where files are processed by RoBERTa/TensorFlow for semantic and structural malware detection. Suspicious attachments enter a sandbox for isolation, followed by sanitization into static previews (e.g., images/text), stored in MySQL, enabling safe user access without execution risks.

3.1 SYSTEM OVERVIEW

Opaline employs a multi-tiered architecture with five key components for proactive email attachment defence.

1. Frontend Tier (User Interface): Flask-based web app with HTML/CSS/JS enables secure attachment uploads and displays sanitized previews.
2. ProcessingTier (Detection): Extracts text viaBeautifulSoup;RoBERTa/TensorFlow analyses semantics and structure for malware/phishing classification.
3. Isolation Tier (Sandbox): Suspicious files execute in a controlled environment to observe behaviour without host compromise.
4. Sanitization Tier: Converts risky attachments to static formats (images/text extracts) for safe viewing, stripping executable code.
5. Backend Tier (Data Layer): MySQL stores scan results/logs; WampServer hosts the full stack for real-time operation.

This layered design ensures detection, containment, and usability.

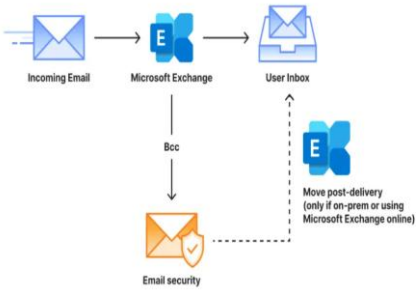

Fig.1 Architecture of the System

3.2 USE CASE ANALYSIS

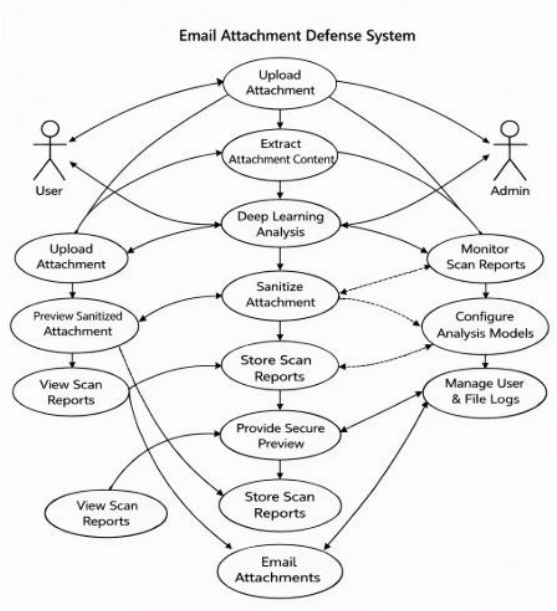

Fig.2 Use Case Diagram

Use Case Flow Example:

1. User uploads attachment: The process begins when a user uploads an email attachment to the system for security analysis.
2. Content extraction: The system extracts textual and structural features from the attachment using parsing and NLP techniques.

3. Deep learning analysis: The extracted data is analysed using a deep learning model to detect malicious patterns or threats.
4. Classification of files: The system classifies the attachment as either safe or suspicious.
5. Safe file handling: If the file is safe, the system generates a secure preview and allows the user to access or download it.
6. Suspicious file handling: Suspicious files are sent to a sandbox environment for further analysis and sanitization.
7. Database storage: All results, logs, and file details are stored in the database for monitoring and future reference.

3.3 CLASS DIAGRAM ANALYSIS

A class diagram represents the static structure of a system by showing the main classes, their relationships, and interactions. In the Opaline Attachment Defence System, the class diagram models components such as users, attachments, detection modules, sandbox environments, and databases. It helps in understanding how different system entities collaborate to detect, analyse, and sanitize malicious email attachments.

Key Components and Their Roles

1. User Class

The User class represents individuals who interact with the system by uploading email attachments for security analysis. Users can access the system, submit files, and view safe previews of processed attachments.

2. Admin Class

The Admin class manages the overall system operations and monitors security activities. Administrators oversee user registrations, review system statistics, and ensure the proper functioning of the attachment detection system.

3. Attachment Class

The Attachment class represents the email files uploaded to the system for analysis. It manages file-related information and acts as the main input for the malicious detection process.

4.Feature Extractor Class

The Feature Extractor class processes uploaded attachments to extract textual and structural information. This extracted data is used to prepare the file for further analysis by the deep learning detection model.

5.Detection Model Class

The Detection Model class uses deep learning techniques to analyse extracted features and identify malicious patterns in attachments. It determines whether the file is safe or potentially harmful.

6.Sandbox Class

The Sandbox class provides a secure and isolated environment to analyse suspicious attachments. It prevents harmful files from affecting the system while enabling deeper inspection.

7. Database Class

The Database class stores system data such as user details, uploaded files, analysis results, and activity logs. It ensures proper record management and supports monitoring and reporting functions within the system.

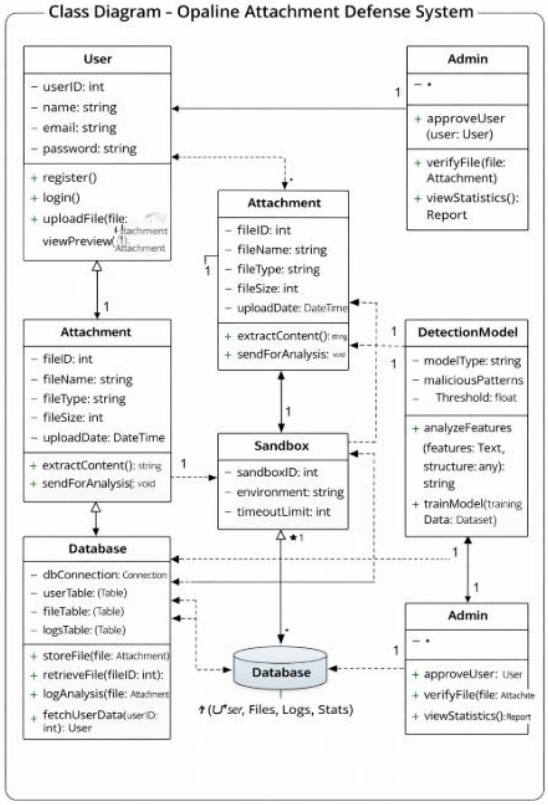

Fig.3 Class Diagram

IV. ATTACHMENT THREAT DETECTION AND SANITIZATION ALGORITHM

The core functionality of the proposed system relies on an intelligent attachment analysis and threat detection mechanism. Unlike conventional antivirus systems that mainly rely on signature-based detection, the proposed platform uses deep learning techniques to analyse the content and structure of email attachments. This approach enables proactive identification of malicious files and provides secure sanitized previews for users.

The attachment detection and sanitization process can be described in the following stages:

4.1 Attachment Data Collection

The system receives email attachments uploaded by users or extracted from incoming emails. Each attachment is stored temporarily and analysed before it is made accessible to the user.

Let:
A represent an attachment file
F_a represent the set of extracted features from the attachment

These features include textual content, metadata, structural properties, and embedded scripts.

4.2 Feature Extraction

Once the attachment is uploaded, the system performs feature extraction to identify meaningful information from the file. This process analyses both textual and structural components of the document.

Let:
T_a represent the textual features extracted from the attachment
S_a represent the structural features of the attachment

These features are extracted using text processing and document parsing techniques.

4.3 Malicious Content Detection

After feature extraction, the system applies a deep learning-based detection model to determine whether the attachment contains malicious content.

Let:
M(A) represent the detection model applied to attachment A

The detection model evaluates patterns in the extracted features and classifies the attachment as either safe or suspicious.

Classification output:

DetectionResult(A) =
{
Safe, if no malicious patterns are detected
Suspicious, if potential malicious behavior is identified
}

This approach improves detection accuracy by identifying hidden threats beyond traditional signature matching.

4.4 Sandbox Analysis

If the attachment is classified as suspicious, the system sends the file to a secure sandbox environment for deeper inspection. The sandbox isolates the attachment and observes its behavior without affecting the main system.

Let:

B(A) represent the behavioral analysis performed in the sandbox.

This stage helps identify hidden malware, scripts, or exploit attempts embedded in the attachment.

4.5 Attachment Sanitization

After analysis, potentially harmful attachments are sanitized before being presented to the user. The sanitization process converts the file into a safe static format, removing any embedded scripts or executable content.

SanitizedFile(A) = SecurePreview(A)

This ensures that users can safely view the attachment without executing the original potentially harmful file.

4.6 *Secure Attachment Delivery*

Finally, the system provides users with secure access to the sanitized attachment preview. All analysis results and file processing logs are stored in the system database for monitoring and future reference.

The system prioritizes user safety while maintaining normal communication workflows by ensuring that only safe or sanitized files are accessible to users.

V. RESULTS AND DISCUSSION

5.1 RESULTS

The proposed system successfully provides a secure and intelligent framework for detecting and sanitizing malicious email attachments. The implementation demonstrates that email attachments can be uploaded and automatically analysed through a structured workflow that includes feature extraction, deep learning–based threat detection, sandbox analysis, and secure preview generation.

The system analyses both textual and structural components of attachments to identify potentially malicious patterns. Using deep learning techniques, the system classifies attachments as either safe or suspicious before they reach the user. Safe attachments are directly converted into secure previews, while suspicious attachments undergo additional inspection within a sandbox environment to prevent system compromise.

The attachment sanitization module further enhances security by converting potentially harmful files into static, non-executable formats. This allows users to safely preview the content without executing embedded scripts or malicious code. The system also stores analysis results and file activity logs in the database for monitoring and future reference.

The integration of automated threat detection and file sanitization significantly reduces the risk of malware infections and phishing attacks caused by malicious email attachments. Overall, the system improves email security while maintaining normal communication workflows and ensuring safe access to shared files.

5.2 Comparison with Existing System

Traditional email security systems primarily rely on signature-based antivirus detection or basic attachment scanning methods. These systems are effective against known malware but often struggle to detect new or sophisticated threats such as zero-day attacks or heavily obfuscated malicious files. In many cases, suspicious attachments may still reach users before being identified, increasing the risk of system compromise.

In contrast, the proposed system introduces a proactive and intelligent detection mechanism using deep learning models to analyse attachment content and structure. This approach enables the system to identify hidden malicious patterns beyond traditional signature matching. Additionally, the use of sandbox environments allows suspicious attachments to be safely executed and analysed in an isolated environment without affecting the main system.

Another significant improvement is the attachment sanitization mechanism, which converts files into secure static previews before user access. This ensures that users can safely view the attachment content without executing potentially harmful code.

Overall, the proposed system enhances email security by combining deep learning–based detection, sandbox analysis, and attachment sanitization. Compared to traditional email security solutions, the system provides improved threat detection capability, safer user interaction with attachments, and stronger protection against emerging cyber threats.

VI. CONCLUSION

The Opaline Attachment Defence System provides an effective solution for detecting and preventing threats originating from malicious email attachments. The system integrates deep learning–based detection, feature extraction, sandbox analysis, and attachment sanitization to identify potentially harmful files before they reach the user. By analysing both textual and structural characteristics of attachments, the system improves the accuracy of threat detection and reduces reliance on traditional signature-based methods. Suspicious files are safely isolated and converted into secure static previews, allowing users to access information without executing malicious content. This approach significantly minimizes the risk of malware infections and phishing attacks through email attachments. Overall, the proposed system enhances email security while maintaining smooth communication workflows. The integration of intelligent detection mechanisms and secure preview generation makes the system a reliable approach for protecting users from evolving email-based cyber threats.

VII. FUTURE SCOPE

The Opaline Attachment Defence System can be further enhanced by integrating advanced artificial intelligence and real-time threat intelligence capabilities. In the future, the system can incorporate more sophisticated deep learning models and larger training datasets to improve the accuracy of malicious attachment detection. Integration with real-time global threat intelligence feeds can help the system quickly identify newly emerging malware patterns and phishing techniques. Additionally, the platform can be extended to support direct integration with popular email services and enterprise email servers to provide real-time protection for incoming attachments. Another potential improvement is the

implementation of behavioral analysis and anomaly detection techniques to identify previously unseen threats more effectively. The system can also be expanded to analyse a wider range of file formats and compressed files. These enhancements will strengthen the system's ability to proactively detect and mitigate evolving cyber threats in email communication environments.

REFERENCES

[1] S. A. Alsuhibany et al., "Analysis and Comparison of BERT and RoBERTa Transformer Models for Email Spam Classification," *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, 2026. (RoBERTa excels in spam recall via semantic understanding.)

[2] A. Mahfoud et al., "Improving malicious email detection through novel ensemble deep learning framework," *Neural Networks*, vol. 154, pp. 397-410, 2022. (Ensemble DL for headers, body, attachments; AUC 0.993.)

[3] N. Vinayakumar et al., "Metadata driven malicious URL detection using RoBERTa large and metadata projection network," *Expert Systems*, 2026. (RoBERTa + metadata for phishing/malware; 98% accuracy.)

[4] J. Smith, "Zero-day malware detection: Leveraging dynamic binary instrumentation and deep learning," *PhD Thesis, Edith Cowan University*, 2025. (DL frameworks like Pulse/Alpha achieve 96-100% on zero-day malware.)

[5] OPSWAT Research Team, "Best Practices for Email Security: Proactive vs. Reactive Approach," OPSWAT Blog, 2024. (Attachment sanitization to neutralize threats pre-delivery.)