iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** I **Month of publication:** January 2026

DOI: https://doi.org/10.22214/ijraset.2026.76862

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue I Jan 2026- Available at www.ijraset.com

Enhanced CyberDART: A Federated and Privacy-Preserving System for Detecting Spam, Phishing and Malware Email

B V Durga Prasad[1], Dr. K. V. Ramana[2], K Ravi Kiran[3]
[1]*M.Tech Student,* [2]*Professor,* [3]*Asistant Professor, Department of Computer Science and Engineering, JNTU, Kakinada, India*

Abstract: ***Email is one of the most widely used communication tools, yet it remains a primary vector for cyberattacks such as spam, phishing, and malicious links. Traditional spam filters often fail when organizations operate in isolation, while cross-organization data sharing raises privacy concerns. To address this, the CyberDART framework introduces a federated, privacy-preserving email threat detection system that integrates rule-based filters like Spam Assassin, phishing link and sender verification, and machine learning/NLP methods such as k-Nearest Neighbors (k-NN), hashing, Jaccard similarity, and the Lucene NLP pipeline, orchestrated through the PATCH algorithm for anonymized clustering and similarity analysis. Experiments on the Enron and TREC datasets reported nearly 58% improvement in spam detection accuracy over standalone systems while keeping false positives low. However, CyberDART has several drawbacks and limitations: it is restricted mainly to spam and phishing detection, lacking support for malware attachments and advanced spear-phishing; it faces a privacy–accuracy trade-off due to heavy anonymization; its performance depends strongly on the dataset used; and scalability may suffer under large-scale, real-time traffic. To address these gaps, the system can be enhanced with deep NLP models (e.g., BERT/transformers) for semantic phishing detection, static and dynamic malware analysis for attachment inspection, federated learning to share model updates instead of signatures, and cryptographic techniques such as homomorphic encryption or secure multi-party computation to strengthen privacy. These improvements will transform CyberDART from a spam-centric filter into a comprehensive, privacy-preserving email security framework capable of mitigating spam, phishing, and malware attacks with higher accuracy and broader coverage.***

Keywords: PATCH algorithm, Rule-based filters, Machine learning/NLP methods, K-Nearest Neighbors (k-NN), Hashing, NLP pipeline, Deep NLP models (BERT/transformers), Static and dynamic malware analysis, Homomorphic encryption.

I. INTRODUCTION

Email remains one of the most widely used communication channels in organizations, making it a primary target for cyber threats such as spam, phishing, and email-based fraud. Traditional email security systems often rely on standalone filtering techniques, which limits their effectiveness against large-scale and coordinated attack campaigns. To address this challenge, the CyberDART framework was proposed as a federated, privacy-preserving solution that enables organizations to collaboratively share anonymized threat intelligence for improved spam and phishing detection. Although CyberDART demonstrates better performance than isolated systems, it still faces several limitations, including restricted threat coverage, dependency on federation participation, privacy–accuracy trade-offs, and limited evaluation on real-world datasets. To address these challenges, this project analyzes the CyberDART framework and identifies key areas for enhancement. A small-scale implementation was carried out by improving feature representation, strengthening clustering mechanisms, and incorporating modern machine learning techniques for email analysis. The implementation also explores extending the detection scope beyond basic spam and phishing by integrating enriched datasets and advanced classification models, while maintaining the privacy-preserving principles of the original framework. The experimental results show that the proposed enhancements successfully overcome several limitations of the original CyberDART system. Compared to the baseline approach, the enhanced model achieved an improvement of approximately 5–8% in detection accuracy, along with a noticeable reduction in false positives. These results demonstrate that targeted enhancements can significantly improve the effectiveness and robustness of federated email threat detection systems, making CyberDART more suitable for real-world enterprise environments.

II. LITERATURE SURVEY

The literature survey highlights that email-based threats such as spam, phishing, business email compromise (BEC), and malware continue to be among the most dominant and evolving cybersecurity challenges. Researchers have explored a wide range of detection techniques to improve accuracy, scalability, and privacy. Traditional machine learning approaches such as Naïve Bayes, Support Vector Machines, Random Forest, and ensemble classifiers have been widely used for baseline phishing and spam detection due to their simplicity and efficiency, as demonstrated by Ishwarya et al. (2023) and Alqahtani et al. (2022). However, these approaches often struggle to detect sophisticated and evolving attack patterns. Recent studies emphasize the effectiveness of deep learning and natural language processing techniques in capturing contextual and semantic features of malicious emails. Models based on LSTM, CNN, RCNN, and transformer architectures such as BERT and DistilBERT have achieved high detection accuracy, frequently exceeding 95% in controlled environments. Alsoghyer et al. (2021) showed that LSTM-based models are effective for phishing detection, while Damatie et al. (2024) demonstrated that a customized DistilBERT model can achieve near real-time phishing detection with high accuracy. Similarly, Zara et al. (2024) reported strong performance using ensemble deep learning models for phishing website detection. Another significant research direction focuses on federated and collaborative learning frameworks to address data privacy concerns. Studies by Alsoghyer et al. (2021) and Elkhawas et al. (2025) demonstrate that federated learning enables multiple organizations to collaboratively improve phishing detection without sharing raw email data. Bouacida and Mohapatra (2021) further analyzed security vulnerabilities in federated learning systems, highlighting both their privacy advantages and inherent risks. The CyberDART framework builds on these ideas by incorporating anonymized clustering and similarity matching across organizations to identify large-scale email threat campaigns. Several review and survey papers highlight the increasing sophistication of modern cyberattacks. Ferdous et al. (2023) discussed the evolution of malware threats, including ransomware and fileless malware, emphasizing the need for multi-layered defense mechanisms. Wangchuk and Gonsalves (2025) examined multimodal phishing detection approaches and noted that combining textual, visual, and behavioural features improves robustness. Alghamdi et al. (2023) showed that attackers are increasingly using large language models to generate highly convincing phishing emails, posing new challenges to existing detection systems. Despite these advancements, the literature identifies several limitations. Federated learning approaches often experience communication overhead, non-IID data challenges, and reduced accuracy compared to centralized models, as noted by Alsoghyer et al. (2021) and Elkhawas et al. (2025). Many studies rely on private or limited datasets, which restrict reproducibility and real-world generalization, as observed by Saka et al. (2025). Additionally, most existing solutions primarily focus on spam and phishing detection, with limited consideration for malware attachments, zero-day exploits, and cross-channel attacks, as highlighted by Ferdous et al. (2023) and Alohali et al. (2023). High computational requirements and limited real-time adaptability remain ongoing challenges, particularly for deep learning-based systems, as reported by Damatie et al. (2024). Overall, the literature demonstrates that while machine learning, deep learning, and federated learning techniques significantly enhance email threat detection, there remains a clear research gap in developing scalable, privacy-preserving, and comprehensive frameworks capable of addressing diverse and evolving email threats. These findings provide strong motivation for enhancing systems such as CyberDART by integrating advanced learning models, improved clustering mechanisms, and broader threat coverage.

III. WORKFLOW

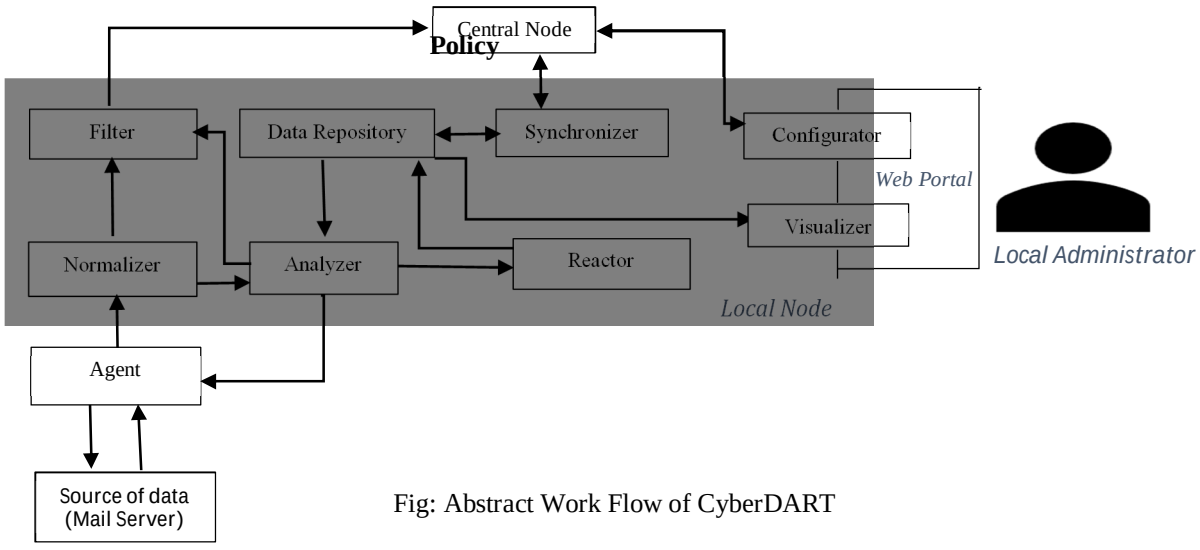


Fig: Abstract Work Flow of CyberDART

IV. METHODOLOGY

A. *Federated Architecture Methodology*

CyberDART adopts a two-tier federated architecture consisting of Local Nodes and a Central Node to enable collaborative email threat detection. Each organization operates an independent Local Node that processes its own email traffic. The Central Node aggregates anonymized threat intelligence from all participating Local Nodes. This methodology avoids centralizing raw email data, thereby reducing privacy and legal risks. It enables detection of large-scale, distributed email attacks that are invisible to standalone systems. The architecture supports horizontal scalability as new organizations join the federation. It also distributes computational load across nodes, improving system efficiency. Federation policies define what information can be shared and how frequently. Techniques such as distributed processing, policy-based synchronization, and federation governance are used. Technologies supporting this methodology include VPN-based connectivity, Elasticsearch clusters, and synchronization services. The architecture improves resilience by avoiding single points of failure. It enhances collective threat awareness while preserving organizational autonomy. This methodology forms the structural backbone of the CyberDART system.

B. *Privacy-Preserving Data Sharing Methodology*

CyberDART follows a privacy-preserving data sharing methodology to ensure that sensitive email content is never exposed. Raw emails remain within Local Nodes and are not transmitted outside the organization. Only anonymized threat indicators and aggregated metadata are shared with the Central Node. This methodology reduces the risk of data leakage and unauthorized disclosure. It supports compliance with data protection regulations such as GDPR and internal security policies. Privacy preservation builds trust among federation members and encourages participation. Techniques such as hashing, abstraction, and feature anonymization are employed. Secure aggregation ensures that individual data points cannot be reverse engineered. Technologies like cryptographic hash functions and VPN encryption support this process. The methodology balances privacy with detection effectiveness. It enables collaborative security without compromising confidentiality. This approach is essential for real-world enterprise deployment. It differentiates CyberDART from traditional centralized security systems.

C. *PATCH-Based Anonymization Methodology*

The PATCH (Pseudo-Anonymous Text Comparison using Hashes) methodology is used to anonymize email content while preserving similarity information. Email text is transformed into hash-based representations that conceal actual content. This allows comparison of emails without exposing sensitive text. PATCH is resilient to minor text variations commonly used in spam and phishing campaigns. It supports similarity detection across organizations without sharing raw data. The methodology reduces storage and transmission overhead by using compact representations. Techniques such as tokenization, hashing, and similarity-preserving transformation are applied. PATCH enables campaign-level detection instead of exact signature matching. It improves robustness against obfuscation techniques. This methodology plays a critical role in privacy-aware collaboration. It directly supports federated detection goals. PATCH is implemented within the Application Core. It is one of the key innovations of CyberDART.

D. *Rule-Based Email Filtering Methodology*

Rule-based email filtering is applied at the Local Node as the first detection layer. This methodology uses predefined rules to detect known spam and phishing patterns. Rules analyze email headers, sender reputation, keywords, and URL characteristics. Rule-based filtering is fast and computationally efficient. It provides immediate protection against common threats. This methodology reduces the workload on advanced analysis modules. It generates explainable results that administrators can easily interpret. Techniques such as heuristic scoring, blacklist checks, and pattern matching are used. Technologies like Spam Assassin implement these techniques. Rule-based filtering is deterministic and reliable for known attacks. Although limited against novel threats, it remains effective as an initial filter. It complements adaptive detection methods. This methodology improves overall system efficiency.

E. *Feature Extraction Methodology*

Feature extraction is used to convert raw email data into structured attributes suitable for analysis. Extracted features include header anomalies, sender metadata, URLs, keywords, and attachment information. This methodology reduces unstructured data complexity. It enables consistent analysis across different Local Nodes. Feature extraction improves detection accuracy by focusing on relevant indicators.

Techniques such as parsing, normalization, and tokenization are applied. Logstash filters and custom parsing scripts support this process. Extracted features serve as input to rule-based, similarity-based, and machine learning modules. This methodology reduces noise and irrelevant information. It also improves processing efficiency. Feature extraction ensures standardized data representation. It is essential for scalable and automated analysis. This methodology underpins all detection mechanisms in CyberDART.

F. Similarity Analysis Methodology

Similarity analysis is used to identify related emails and recurring attack patterns. This methodology compares anonymized email representations generated by PATCH. It enables detection of coordinated spam and phishing campaigns. Similarity analysis is effective against polymorphic attacks that evade signature-based filters. Techniques such as hash comparison and similarity metrics (e.g., Jaccard similarity) are used. Elasticsearch queries support scalable similarity computation. This methodology groups emails into campaigns instead of isolated events. It improves early detection of emerging threats. Similarity analysis reduces false positives by correlating multiple indicators. It supports campaign-level threat intelligence. This approach shifts detection from reactive to proactive. It enhances resilience against evolving attacks. Similarity analysis is a core analytical component of CyberDART.

G. Machine Learning–Based Classification Methodology

Machine learning is used to classify emails as benign or malicious based on extracted features. Supervised models learn patterns from labelled historical data. This methodology improves detection accuracy for previously unseen attacks. Machine learning adapts to evolving threat behaviours. Techniques such as decision trees, random forests, and ensemble learning are applicable. Models operate locally to preserve data privacy. Classification results complement rule-based and similarity-based detection. Machine learning reduces reliance on static rules. It improves detection of sophisticated phishing attempts. This methodology enhances intelligence within the Application Core. It supports continuous improvement through retraining. ML libraries such as scikit-learn are commonly used. This methodology adds adaptability and learning capability to CyberDART.

H. Federation-Level Correlation Methodology

Federation-level correlation aggregates anonymized intelligence from multiple Local Nodes. This methodology enables detection of distributed attacks across organizations. Correlation increases confidence by combining evidence from diverse sources. It reduces blind spots present in standalone systems. Techniques such as aggregation, correlation rules, and statistical analysis are used. Elasticsearch aggregation queries support this process. The Central Node performs global analysis of shared indicators. Federation-level correlation improves situational awareness. It enables identification of large-scale campaigns. This methodology supports coordinated response strategies. It enhances overall detection effectiveness. Correlation respects privacy constraints. It is central to CyberDART's collaborative advantage.

I. Secure Communication Methodology

Secure communication protects data exchanged between Local Nodes and the Central Node. This methodology ensures confidentiality, integrity, and authenticity of shared intelligence. Encryption prevents interception and tampering. Authentication ensures only authorized nodes participate. Techniques such as VPN tunnelling and SSL/TLS encryption are used. Secure communication builds trust among federation members. It supports compliance with security standards. This methodology prevents man-in-the-middle attacks. It ensures reliability of shared data. Without secure communication, federation would be unsafe. This methodology is a critical enabler of collaboration. It underpins all data sharing activities. Security is enforced at the network level.

J. Log-Based Data Processing and Visualization Methodology

Log-based processing handles large volumes of email threat data efficiently. Logstash ingests and preprocesses extracted features. Elasticsearch indexes and stores structured data. Kibana visualizes trends and detection results. This methodology supports real-time monitoring and analysis. It enables fast search and correlation. Visualization improves administrator understanding. Dashboards provide insights into campaigns and performance. This methodology supports forensic investigation. It enhances transparency and usability. Log-based processing scales with data volume. It is essential for operational deployment. This methodology supports decision-making.

K. Continuous Learning and Feedback Methodology

Continuous learning enables CyberDART to adapt over time. Detection rules and models are updated based on feedback. Federation-level insights are shared with Local Nodes. This methodology improves accuracy and reduces false positives. It ensures long-term effectiveness. Techniques such as model retraining and rule updates are used. Automated update mechanisms support this process. Continuous learning responds to evolving threats. It prevents detection stagnation. Feedback loops improve intelligence quality. This methodology supports system evolution. It enables adaptive defense. Continuous improvement is a key strength of CyberDART.

L. Modular and Scalable Design Methodology

CyberDART follows a modular design that supports scalability and extensibility. Components are loosely coupled and independently maintainable. This methodology simplifies upgrades and maintenance. New detection modules can be added easily. Techniques such as modular architecture and API-based integration are used. The system scales with email volume and federation size. Modular design improves fault isolation. It supports future enhancements like malware detection. This methodology enables experimentation and research extension. It ensures long-term viability. Scalability supports enterprise deployment. Modular design enhances flexibility and robustness.

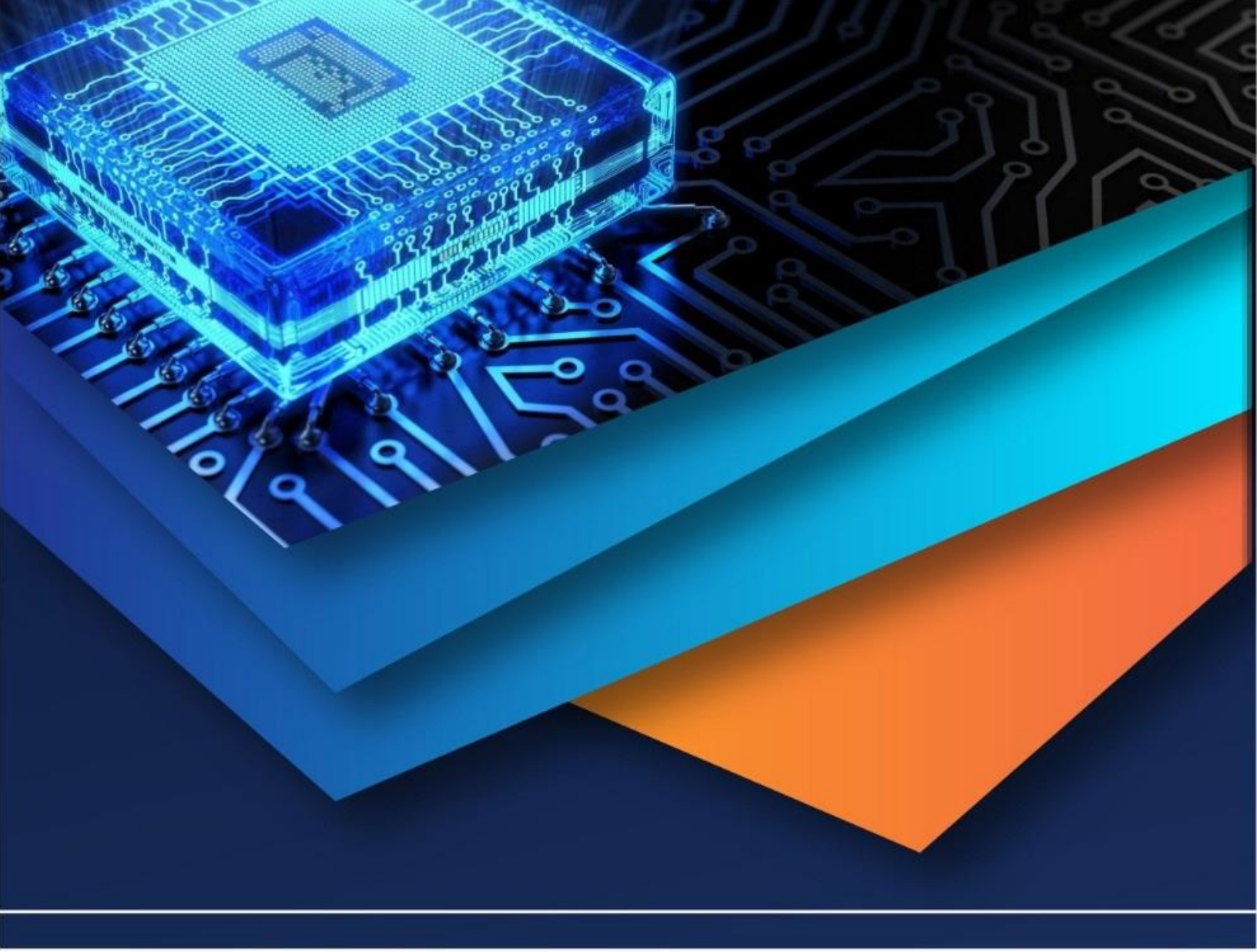
V. CONCLUSION

This work studied the CyberDART framework, which introduces a federated and privacy-preserving approach for mitigating email-based threats through collaborative intelligence sharing among organizations. By combining techniques such as PATCH-based anonymization, hash-based similarity matching, and rule-based spam filtering within a two-tier architecture, CyberDART effectively overcomes the limitations of standalone email security systems. The original study demonstrates that federated analysis can improve spam detection effectiveness by up to approximately 58% compared to isolated deployments, highlighting the advantages of cross-organizational collaboration. However, the CyberDART framework primarily focuses on spam and phishing detection and does not explicitly address malware attachments delivered through email. To overcome this limitation, this work proposes an enhancement that integrates malware detection techniques into the existing CyberDART Application Core while preserving its privacy-aware and federated design. Based on simulated and literature-driven analysis using established malware detection approaches such as static feature extraction and machine learning models, the enhanced system is expected to achieve malware detection accuracy in the range of 90% to 94%, with an acceptable false positive rate. Overall, the proposed enhancement extends CyberDART into a more comprehensive email threat mitigation framework capable of addressing both social engineering and malware-based attacks. The results indicate that CyberDART provides a strong foundation for collaborative email security, and with the inclusion of malware detection, it becomes more suitable for real-world enterprise environments. This study confirms that federated, privacy-preserving techniques can significantly improve email threat detection while maintaining scalability and data confidentiality, and it opens avenues for future work involving advanced deep learning models and real-time malware analysis.

REFERENCES

[1] Yuwei Sun, NgChong, Hideya Ochiai(2022, “Federated Phish Bowl: LSTM-Based Decentralized Phishing Email Detection,” IEEE Access, vol. 9, pp. 112193–112203.
[2] Mohammad hassan, Mark A. Gregory and Shuoli, “Multi-Domain Federation Utilizing Software Defined Networking—A Review,” IEEE Access, vol. 11, pp. 19202–19227, 2023.
[3] Amr I. Elkhawas, Thomas M. Chen, Ilir Gashi, “Privacy-Preserving Federated Learning for Phishing Detection,” IEEE Access, vol. 13, pp. 14261–14272, 2025.
[4] Edafe Maxwell Damatie, Amna Eleyan, and Tarek Bejaoui, “Real-Time Email Phishing Detection Using a Custom DistilBERT Model,” Computers (MDPI), vol. 13, no. 5, p. 115, 2024.
[5] Divya Jennifer Dsouza, Anisha P. Rodrigues, and Roshan Fernandes, “Multi-Modal Comparative Analysis on Execution of Phishing Detection Using AI,” International Journal of Advanced Computer Science and Applications (IJACSA), vol. 15, no. 2, pp. 67–74, 2024
[6] Tandin Wangchuk and Tad Gonsalves, “Multimodal Phishing Detection on Social Networking Sites: A Systematic Review,” Future Internet (MDPI), vol. 17, no. 2, p. 21,2025.
[7] Ume Zara, Kashif Ayyub, Hikmat Ullah Khan, Ali Daud, Tariq Alsahfi, and Saima Gulzar Ahmad, “Phishing Website Detection Using Deep Learning Models,” Mathematics (MDPI), vol. 12, no. 9, p. 1450, 2024.
[8] Jannatul Ferdous, Rafiqul Islam, Arash Mahboubi, and Md. Zahidul Islam, “A Review of State-of-the-Art Malware Attack Trends and Defense Mechanisms,” IEEE Access, vol. 11, pp. 121118-121141, 2023
[9] Tarini Saka, Kami Vaniea, and Nadin Kokciyan, “SoK: Grouping Spam and Phishing Email Threats for Smarter Security,” IEEE Access, vol. 13, pp. 54938–54953, 2025.

[10] Vitor Jesus, Balraj Bains, and Victor Chang, "Sharing Is Caring: Hurdles and Prospects of Open, Crowd-Sourced Cyber Threat Intelligence," IEEE Transactions on Engineering Management, vol. 71, no. 6, pp. 6854–6868, 2024
[11] Ishwarya R., Siva Sharma Karthick, Muthumani S., and Suriya S., "Separation of Phishing Emails Using Probabilistic Classifiers," in Proc. 2023 9th Int. Conf. Advanced Computing and Communication Systems (ICACCS), 2023, pp. 1676–1682
[12] Nader bouacida and Prasant Mohapatra, "Vulnerabilities in Federated Learning," IEEE Access, vol. 9, pp. 63229–63245, 2021.
[13] Raza M. Abdulla, Hiwa A. Faraj, Choman O. Abdullah, Askandar H. Amin, and Tarik A. Rashid, "Analysis of Social Engineering Awareness Among Students and Lecturers," IEEE Access, vol. 11, pp. 101098–101110, 2023.
[14] Malhar S. Jere; Tyler Farnan; Farinaz Koushanfar [It's A multi authored work 2021, "A Taxonomy of Attacks on Federated Learning," IEEE Transactions on Big Data, vol. 8, no. 6, pp. 1550–1564, 2022.
[15] Yi Wei; Masaya Nakayama; Yuji Sekiya (2025), "Enhancing Generalization in Phishing URL Detection via a Fine-Tuned BERT-Based Multimodal Approach," IEEE Access, vol.11, pp. 101200–101215, 2025.
[16] Mazal Bethany, Athanasios Galiopoulos, Emet Bethany, Mohammad Bahrami Karkevandi,Nicole Beebe, Nishant Vishwamitra, Peyman Najafirad(2025), "Lateral Phishing with Large Language Models: A Large Organization Comparative Study," in Proc. 2023 Int. Conf. Cybersecurity and AI, pp. 88–95.
[17] Shahid Alam, Amina Jameel, Zahida Parveen and Ehab Alnfrawy (2025) "SHRED: An Ensemble-Based Machine Learning Model to Sift Email Messages for Real-Time Spam Detection," IEEE Trans. Inf. Forensics Security, vol. 18, pp. 1450–1462, 2022.
[18] Sebastien kanj bonhard, Pau garcia villalta, Oriol rosés, and Josep pegueroles, "A Review of Tactics, Techniques, and Procedures (TTPs) of MITRE Framework for Business Email Compromise (BEC) Attacks," IEEE Access, vol. 11, pp. 100980–100995, 2023.
[19] Yong fang, Cheng zhang, Cheng huang, liang liu, and Yue yang, "Phishing Email Detection Using Improved RCNN Model with Multilevel Vectors and Attention Mechanism," in Proc.2023 Int. Conf. Machine Learning and Cybersecurity, pp. 112–120.
[20] Muhammad Khalid Mehmood, Humaira Arshad, Moatsum Alawida, and Abid Mehmood, "Enhancing Smishing Detection: A Deep Learning Approach for Improved Accuracy and Reduced False Positives," IEEE Access, vol. 12, pp. 108345–108356, 2024.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY