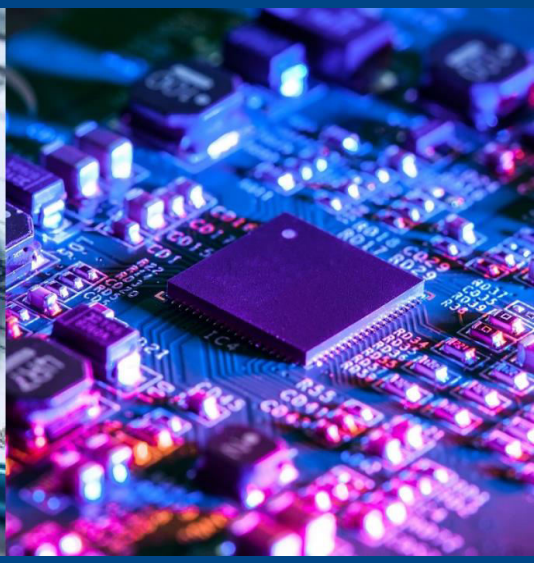
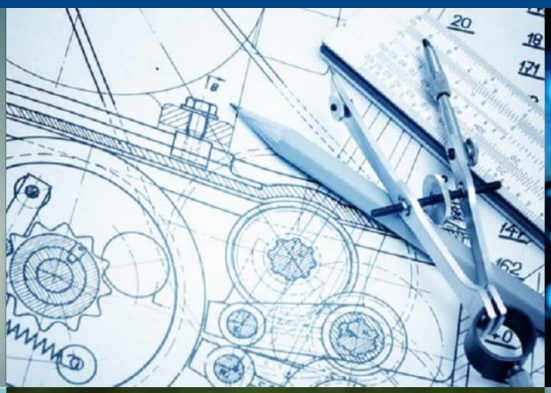
ISSN: 2582-7219

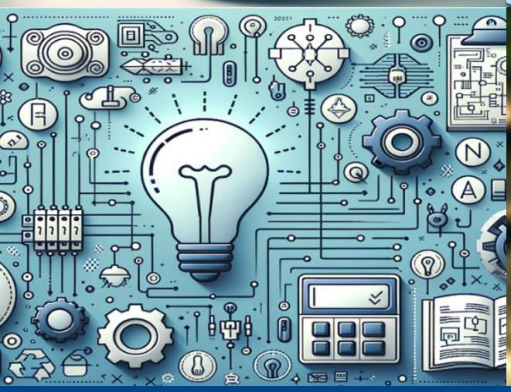




International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206 **Volume 9, Issue 4, April 2026**

© 2026 IJMRSET | Volume 9, Issue 4, April 2026| DOI:10.15680/IJMRSET.2026.0904470

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Ethical Email Tracking with Integrated Web Scrapping for Real-Time Policy Monitoring

Sandip Gavali, Onkar Ingale, Pratiksha Dhere, Pooja Muske, Tanishka Karande, Shivani Malve, Dr. P.A. Satarkar

Student, Dept. of CSE, SVERI's COE, Pandharpur Solapur, India

Student, Dept. of CSE, SVERI's COE, Pandharpur Solapur, India

Student, Dept. of CSE, SVERI's COE, Pandharpur Solapur, India

Student, Dept. of CSE, SVERI's COE, Pandharpur Solapur, India

Student, Dept. of CSE, SVERI's COE, Pandharpur Solapur, India

Student, Dept. of CSE, SVERI's COE, Pandharpur Solapur, India

HOD, Dept. of CSE, SVERI's COE, Pandharpur Solapur, India

ABSTRACT: An advanced approach to email security through the development of an Ethical Email Tracking System integrated with Web Scraping for Cyber Threat Detection. In today's digital environment, emails are increasingly used not only for communication but also for phishing attacks, spoofing, malware distribution, and hidden tracking of user behavior. Existing systems primarily focus on spam filtering and lack transparency regarding tracking mechanisms. The proposed system introduces a privacy-aware solution that analyzes email content, headers, attachments, and embedded links to detect malicious activities. It incorporates ethical tracking detection to identify hidden tracking pixels and unauthorized monitoring techniques. Additionally, web scraping is used to safely analyze external links without user interaction. A risk scoring engine classifies emails based on multiple threat parameters. The system ensures ethical compliance by enforcing privacy policies and user consent.

I. INTRODUCTION

Email communication has become an essential part of personal and professional environments. However, it has also become a major vector for cyber threats such as phishing, spoofing, malware attacks, and unauthorized user tracking. Attackers exploit email systems to deceive users into revealing sensitive information or executing malicious actions. Traditional email security systems focus primarily on spam detection and blacklist filtering. While these techniques are effective to some extent, they fail to address modern threats such as hidden tracking pixels, link-based phishing attacks, and spoofed identities. Moreover, there is a lack of transparency in how user data is monitored through email tracking mechanisms.

This project proposes an Ethical Email Tracking System that combines email analysis with controlled web scraping to detect both internal and external threats. The system emphasizes ethical practices by ensuring user privacy, consent, and compliance with security standards. The objective is to provide a comprehensive, transparent, and secure email analysis solution. Machine.

© 2026 IJMRSET | Volume 9, Issue 4, April 2026| DOI:10.15680/IJMRSET.2026.0904470

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

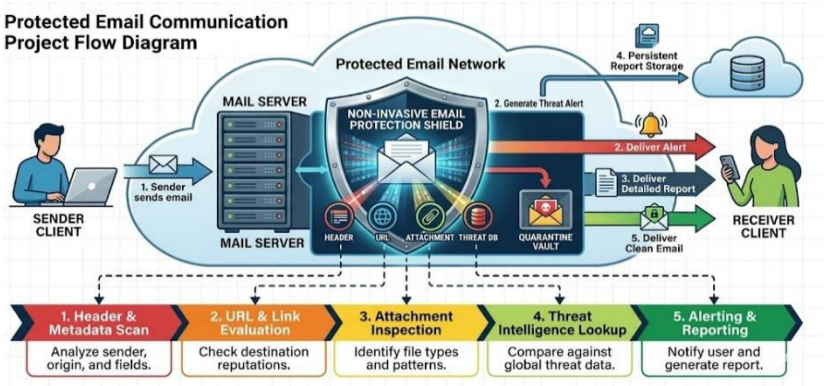


Figure 1: Flow Structure

II. LITERATURE REVIEW

A Several studies have focused on email security, particularly in spam detection and phishing identification. Machine learning-based spam filters have been widely used to classify emails based on textual content and metadata. Similarly, phishing detection systems analyze URLs and domain characteristics to identify malicious links.

Email tracking technologies, such as tracking pixels and unique identifiers, are commonly used in marketing to monitor user engagement. However, these techniques often operate without user awareness, raising privacy concerns. Existing research has highlighted the risks associated with such tracking but lacks integrated solutions to detect and prevent it.

Various researchers have explored email security from different perspectives such as phishing detection, spam filtering, and user behavior analysis.

Sharon et al. proposed a phishing email detection system using machine learning techniques [1].
The study highlights the use of classifiers such as Naive Bayes and Support Vector Machines to identify malicious emails based on content and metadata. However, the approach mainly focuses on classification accuracy and does not consider external threat sources such as malicious links or tracking mechanisms.

Kumar et al. proposed a comprehensive survey on phishing detection techniques [2].
The study explains how attackers exploit user trust through deceptive emails and fake domains. It emphasizes the importance of URL analysis and domain verification. Although effective in identifying phishing attacks, the system lacks integration with real-time monitoring and ethical transparency features.

Rajasekhar et al. proposed a data-driven approach for phishing detection using email header analysis [3]. The technique analyzes inconsistencies in sender information, IP addresses, and mail routing paths to detect spoofing attempts. While this approach improves detection of forged emails, it does not address issues related to hidden tracking or user privacy.

Research conducted by MDPI introduced deep learning-based phishing detection using neural networks such as CNN and LSTM [4].
These models analyze complex patterns in email content and achieve high detection accuracy. However, the system requires large datasets and computational resources, and it does not incorporate ethical compliance or transparency in tracking activities.

Another study on email tracking technologies highlights the use of tracking pixels and web bugs embedded in HTML emails [5].
These techniques allow senders to monitor user behavior such as email open rates and link clicks without user awareness. While this raises significant privacy concerns, existing systems do not provide effective mechanisms to detect or prevent such tracking.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

User behavior analysis has also been studied in the context of phishing detection. Research shows that users often fail to identify phishing emails due to lack of awareness and misleading content [6].
This indicates the need for automated systems that assist users in detecting threats and improving decision-making.

Relevance to current Research
The work presented in this paper takes due care of the data which is sent on the email as it not only provides the integrity check but also security for the data as well. This lets us to test the integrity at the moment we review an email.

No.	Paper Title	Author Name	Key Points	Remark
1	Phishing Email Detection using Machine Learning Techniques	Sharon Abraham et al., 2022	Uses ML algorithms like Naive Bayes and SVM to classify phishing emails based on content and metadata [1]	Improves detection accuracy but lacks external link and tracking analysis
2	A Survey on Phishing Detection Techniques	Santosh Kumar et al., 2021	Explains various phishing detection methods including URL analysis and domain verification [2]	Provides overview but does not include ethical transparency or tracking detection
3	Data-Driven Phishing Detection using Email Headers	Rajasekhar Kalamata et al., 2025	Detects phishing using inconsistencies in email headers, IP addresses, and routing paths [3]	Strong header analysis but ignores tracking pixels and external threats
4	Deep Learning-Based Phishing Detection using CNN and LSTM	MDPI Research, 2023	Uses deep learning models to detect phishing emails with high accuracy [4]	High accuracy but requires large datasets and lacks ethical compliance features
5	Analysis of Email Tracking and Web Bugs in HTML Emails	ScienceDirect Research, 2018	Explains how tracking pixels and web bugs monitor user behavior in emails [5]	Highlights privacy issues but does not provide detection/prevention mechanism

III. METHODOLOGY OF PROPOSED SURVEY

The proposed system follows a **multi-layer pipeline architecture** to analyze emails securely and ethically. The complete methodology is divided into the following steps:

Secure Email Acquisition:
Emails are fetched using Gmail API / IMAP with OAuth-based authentication.
Only read-only access is granted to ensure privacy and security.
Email Parsing & Feature Extraction:
The fetched email is processed using parsing modules.
Extracted components includes Header fields (From, IP, routing path), Email body content, Embedded URLs, Attachments, Hidden HTML elements.

Header Authentication & Spoofing Detection:
The system validates email authenticity using SPF (Sender Policy Framework), DKIM (DomainKeys Identified Mail), DMARC (Domain-based Message Authentication).

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Phishing Detection & Web Scraping Analysis:
Regeneration All extracted URLs are analyzed using Pattern recognition (suspicious keywords, URL length), Domain verification techniques.
Web scraping is applied to Retrieve webpage metadata (title, forms, redirects), Detect fake login pages, Identify malicious redirections

Malware & Attachment Scanning:
Attachments are scanned using antivirus engines like Clam AV.

Tracking Detection & Privacy Analysis:
The system identifies hidden tracking mechanisms like Tracking pixels (1×1 images), Embedded external resources, Unique tracking IDs in URLs.

Risk Scoring & Ethical Compliance:
All module outputs are combined in a Risk Scoring Engine. Each threat is assigned weight and final risk score is calculated

Alert Generation & Reporting:
The system generates Warning alerts for users & Visual risk indicators

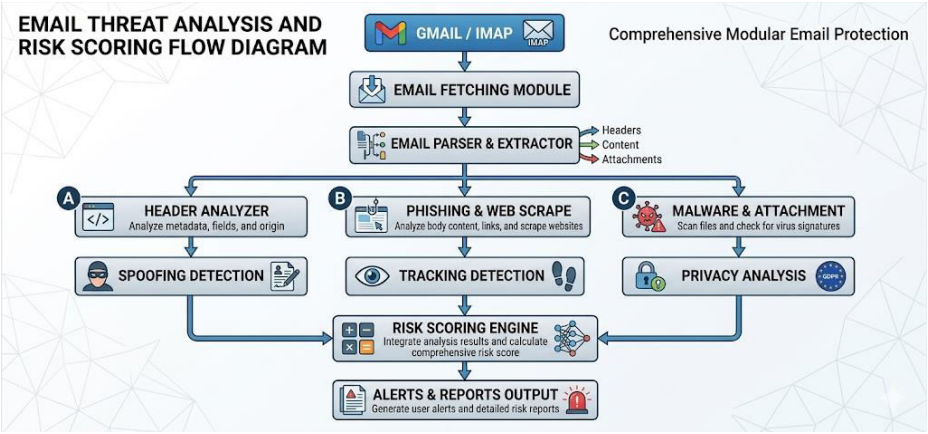


Figure 2: Methodology

IV. CONCLUSION AND FUTURE WORK

This project presents a comprehensive solution for email security by integrating ethical email tracking with web scraping and multi-layer threat detection. The system effectively identifies phishing, spoofing, malware, and tracking activities while maintaining user privacy and transparency. The proposed approach is scalable and adaptable for both individual users and organizational environments. Its modular design allows easy integration with existing email systems and cybersecurity frameworks. Future enhancements may include the use of machine learning for advanced threat detection, real-time deployment in enterprise environments, and integration with cloud-based security platforms. The system has strong potential to contribute to the development of secure and ethical digital communication systems.

REFERENCES

[1] Sharon Abraham, “Phishing Email Detection using Machine Learning Techniques,” IJERT, 2022.
[2] Santosh Kumar, “A Survey on Phishing Detection Techniques,” 2021.
[3] Rajasekhar Kalamata, “Data-Driven Phishing Detection using Email Headers,” 2025.
[4] MDPI Research, “Deep Learning-Based Phishing Detection using CNN and LSTM,” 2023.
[5] ScienceDirect, “Analysis of Email Tracking and Web Bugs in HTML Emails,” 2018.
[6] Frontiers Research, “User Behavior Analysis in Phishing Email Detection,” 2020.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com