International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

PhishNet : AI Powered Email Phishing Detection

Prof. Supriya Chougule
Professor
PDEA's College of Engineerin g Manjari (Bk.),
Pune Pune, Maharashtra, India

Piyush More
Dept. Of Computer Engineering
PDEA's College of Engineering
Savitribai Phule Pune University
Pune, Maharashtra, India

Amar Danwale
Dept. Of Computer Engineering
PDEA's College of Engineering
Savitribai Phule Pune University
Pune, Maharashtra, India

Rushikesh Phalke
Dept. Of Computer Engineering
PDEA's College of Engineering
Savitribai Phule Pune University
Pune, Maharashtra, India

Pratik Pawer
Dept. Of Computer Engineering
PDEA's College of Engineering
Savitribai Phule Pune University
Pune, Maharashtra, India

Abstract **- Email-based phishing remains a major cybersecurity threat, enabling credential theft, Business Email Compromise (BEC), and malware distribution. Traditional email security solutions often struggle against modern phishing techniques such as brand impersonation, URL obfuscation, homoglyph domains, and AI-generated phishing content. This paper presents PhishNet, a modular phishing detection platform that employs a parallel five-node analysis architecture consisting of Sender, Authentication, Content, Link, and Attachment Analysis modules. The system also incorporates an Adversarial Risk Engine for detecting evasion techniques and a Policy Engine for automated response actions.**

PhishNet supports multiple deployment modes, including IMAP-based monitoring, Gmail OAuth analysis, and API-driven email submission. Experimental evaluation on 45 adversarial phishing samples across 15 attack categories achieved a precision of 97.8%, recall of 95.6%, and F1-score of 96.7%. Results demonstrate that the multi-node architecture provides improved detection accuracy and greater resilience against advanced phishing attacks compared to individual analysis methods. The platform offers a scalable and privacy-aware approach for enterprise email security.

Keywords - ***email security; phishing detection; adversarial evasion; weighted scoring; multi-tenant architecture; GDPR compliance; threat intelligence; Shannon entropy; homoglyph detection; enterprise policy engine***

INTRODUCTION

Email phishing has evolved from easily-detectable bulk spam into precision-targeted, context-aware social engineering campaigns. Adversaries now leverage large language models to generate grammatically flawless impersonation content, register look-alike domains days before attacks [13], and encode malicious payloads in Base64 MIME blocks to evade link scanners. According to industry reports, phishing accounts for over 80% of reported security incidents, with credential harvesting and BEC variants causing billions in annual losses.

ML classifiers are evaded by feature manipulation — an attacker who knows a content classifier relies on keyword presence can craft grammatically legitimate text that triggers no alert while embedding the phishing payload in a tracked redirect. Third, privacy regulation GDPR prohibit the bulk ingestion and storage of raw corporate email content that most commercial security gateways perform by default.

This paper addresses these three gaps through PhishNet, which contributes: (1) a five-node parallel analysis pipeline with dynamic reliability-weighted aggregation that prevents single-node evasion;

(2) an independent Adversarial Risk Engine that detects cross-node evasion patterns invisible to any individual analyser; (3) three flexible ingestion modes satisfying diverse enterprise privacy and deployment requirements; and (4) a GDPR-compliant privacy middleware stack performing PII redaction before any data is persisted.

The remainder of this paper is organized as follows. Section II presents the research contributions of PhishNet. Section III defines the threat model, including protected assets, adversary profiles, attack surfaces, and threat assumptions. Section IV describes the overall system model. Section V details the PhishNet system architecture, while Section VI explains the detection methodology and analysis workflow. Section VII presents the security architecture and operating modes. Section VIII introduces the mathematical model used for threat scoring and adversarial assessment. Section IX provides the experimental evaluation and performance analysis. Section X discusses future enhancements and research directions. Finally, Section XI concludes the paper, and Section XII lists the references.

I. RESEARCH CONTRIBUTIONS

Modern phishing attacks increasingly evade traditional email security solutions through techniques such as sender impersonation, authentication bypass, malicious redirects, and social engineering. To address these challenges, PhishNet introduces a multi-layered phishing detection platform that

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

emphasizes explainability, adversarial resilience, and privacy-aware deployment.

Five-Node Parallel Analysis Architecture

PhishNet employs five independent analysis modules Sender, Authentication, Content, Link, and Attachment Analysis to evaluate different attack surfaces and generate specialized threat scores, improving detection robustness.

2.Reliability-Weighted Threat Aggregation

Threat scores are combined using a reliability-weighted framework optimized through Bayesian methods, ensuring that high-confidence indicators have greater influence on the final verdict.

3. Independent Adversarial Risk Engine

A dedicated engine detects advanced evasion techniques, including Unicode homoglyphs, obfuscated content, entropy anomalies, and brand impersonation, while correlating signals across analysis nodes.

4.Multi-Mode Email Ingestion Framework

PhishNet supports IMAP Bulk Analysis, Gmail OAuth On-Demand Analysis, and Email Forward API integration, enabling deployment across personal, organizational, and enterprise environments.

5.Privacy-First Gmail OAuth Analysis

Emails are analyzed only with explicit user consent, and sensitive data can be processed without persistent storage, supporting privacy-focused and compliance-oriented deployments.

6. Explainable Phishing Verdict Generation

PhishNet provides transparent verdicts by reporting analyzer scores, detected indicators, threat intelligence results, and adversarial findings, improving trust and investigation efficiency.

II. THREAT MODEL

A formal threat model grounds the design of PhishNet's detection and mitigation capabilities. This section defines the assets under protection, adversary profiles, attack surface, assumed threat techniques, and system assumptions.

A. Protected Assets

1. Corporate Email Accounts:

Corporate email accounts represent the primary assets protected by PhishNet. Attackers frequently target employee and executive mailboxes through phishing and impersonation campaigns to gain unauthorized access. If compromised, these accounts can facilitate credential theft, business email compromise (BEC), and lateral movement within the organization.

2. User Credentials:

User credentials, including usernames, passwords, and multi-factor authentication codes, are common targets of phishing attacks. PhishNet protects these identity assets by detecting credential-harvesting attempts embedded within malicious emails. A successful compromise could lead to account takeover and large-scale data breaches.

3. OAuth Access Tokens:

OAuth access tokens are used to securely access Gmail data in On-Demand Mode. These short-lived tokens are stored using encrypted mechanisms and provide temporary authorization for email analysis. If compromised, attackers could gain unauthorized access to user inboxes and sensitive email content.

4. Corporate Data and Intellectual Property:

Business communications, contracts, financial records, and proprietary information stored within email systems constitute high-value information assets. A compromised email account may expose confidential organizational data, resulting in intellectual property theft, financial loss, or regulatory compliance violations

5. Email Infrastructure Reputation:

The reputation of organizational email infrastructure, including SPF, DKIM, and DMARC configurations as well as sender IP reputation, is critical for maintaining trusted email communication. If these assets are abused or compromised, the organization's domain may be blacklisted, significantly impacting email deliverability and business operations.

B. Adversary Profiles

1. Phishing-as-a-Service (PhaaS)

Operators:Cybercriminals who use phishing kits, purchased domains, and automated infrastructure to conduct large-scale phishing campaigns.

Business 2. Email Compromise (BEC) Actors:

Attackers who impersonate executives, vendors, or trusted contacts to facilitate financial fraud and sensitive information theft.

3. Advanced Persistent Threat (APT) Actors:

Highly sophisticated adversaries that use advanced phishing techniques for espionage, intelligence gathering, and intellectual property theft.

4.Credential Harvesters:

Attackers focused on stealing usernames, passwords, and authentication tokens through mass phishing campaigns.

5. Insider Threat Actors:

Authorized users who intentionally misuse organizational resources to steal data or bypass security controls.

C. Attack Surface

1. Email Header Manipulation:

Spoofed sender information used to impersonate trusted organizations.

2. Authentication Abuse:

Exploitation of weak or misconfigured SPF, DKIM, and DMARC policies.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

3. Domain and URL Obfuscation:

Use of homoglyph domains, redirect chains, and cloaked URLs to evadedetection.

4. Attachment-Based Malware Delivery:

Malicious files disguised through deceptive naming and double-extension techniques.

5. Content Obfuscation and Social Engineering:

LLM-generated phishing content and encoding techniques designed to bypass filters.

6. OAuth Token Abuse:

Unauthorized access obtained through theft or misuse of authentication tokens.

D. Threat Assumptions

PhishNet assumes that adversaries can generate sophisticated phishing content, abuse legitimate email authentication mechanisms, and rapidly deploy new phishing infrastructure before reputation services can identify it. The system also assumes that external threat-intelligence services may become temporarily unavailable and that phishing destinations may change after email delivery through redirect-based cloaking. Finally, OAuth tokens are assumed to be short-lived and scope-restricted, while internal scoring logic remains inaccessible to attackers.

III. SYSTEM MODEL

The PhishNet system is designed as a multi-layered phishing detection framework that analyzes emails through multiple independent security perspectives before generating a final verdict. Unlike conventional phishing detection systems that rely on a single classifier or isolated feature set, PhishNet employs a collaborative analysis model in which multiple specialized analyzers contribute evidence toward a unified risk assessment.

Email Infrastructure Reputation:

The reputation of organizational email infrastructure, including SPF, DKIM, and DMARC configurations as well as sender IP reputation, is critical for maintaining trusted email communication. If these assets are abused or compromised, the organization's domain may be blacklisted, significantly impacting email deliverability and business operations.

B. Adversary Profiles1. Phishing-as-a-Service (PhaaS)

Operators:Cybercriminals who use phishing kits, purchased domains, and automated infrastructure to conduct large-scale phishing campaigns.

Business 2. Email Compromise (BEC) Actors:

Attackers who impersonate executives, vendors, or trusted contacts to facilitate financial fraud and sensitive information theft.

3. Advanced Persistent Threat (APT) Actors:

Highly sophisticated adversaries that use advanced phishing techniques for espionage, intelligence gathering, and intellectual property theft.

4.Credential

Harvesters:Attackers focused on stealing usernames, passwords, and authentication tokens through mass phishing campaigns.

5. Insider Threat Actors:

Authorized users who intentionally misuse organizational resources to steal data or bypass security controls.

C. Attack Surface

1. Email Header Manipulation:

Spoofed sender information used to impersonate trusted organizations.

2. Authentication Abuse:

Exploitation of weak or misconfigured SPF, DKIM, and DMARC policies.

3. Domain and URL Obfuscation:

Use of homoglyph domains, redirect chains, and cloaked URLs to evadedetection.

4. Attachment-Based Malware Delivery:

Malicious files disguised through deceptive naming and double-extension techniques.

5. Content Obfuscation and Social Engineering:

LLM-generated phishing content and encoding techniques designed to bypass filters.

6. OAuth Token Abuse:

Unauthorized access obtained through theft or misuse of authentication tokens.

D. Threat Assumptions

PhishNet assumes that adversaries can generate sophisticated phishing content, abuse legitimate email authentication mechanisms, and rapidly deploy new phishing infrastructure before reputation services can identify it. The system also assumes that external threat-intelligence services may become temporarily unavailable and that phishing destinations may change after email delivery through redirect-based cloaking. Finally, OAuth tokens are assumed to be short-lived and scope-restricted, while internal scoring logic remains inaccessible to attackers.

IV.SYSTEM ARCHITECTURE

PhishNet follows a modular, multi-layered architecture designed to provide accurate, explainable, and resilient phishing detection. The system combines specialized analysis pipelines for sender authentication, content inspection, URL evaluation, attachment analysis and threat intelligence correlation. Each module independently evaluates a specific attack surface and contributes its findings to a centralized threat aggregation engine, enabling comprehensive detection while maintaining scalability and explainability.

A. Overall System Architecture

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

The PhishNet architecture explains how the system detects phishing emails in a structured step-by-step process. First, the email enters through the Client Layer and is sent to the FastAPI API Gateway, which performs authentication, request validation, and traffic control to ensure secure and reliable processing. This gateway acts as the entry point of the system.

Next, the Analysis Orchestrator handles the email by distributing it to multiple analysis modules. Each module focuses on a specific security aspect:

- URL Analysis checks embedded links for malicious websites, suspicious redirects, and risky domains.
- Sender Authentication Analysis verifies if the sender is genuine using SPF, DKIM, and DMARC protocols.
- Content Analysis scans the email text and attachments for phishing keywords, urgent language, or suspicious patterns.
- Threat Intelligence Integration checks external threat databases like VirusTotal and AbuseIPDB for known malicious indicators.

The outputs from all modules are sent to the Deterministic Threat Aggregator, where they are combined using predefined rules and weighted scoring. This component calculates the final threat score and classifies the email as Safe, Suspicious, or Phishing.

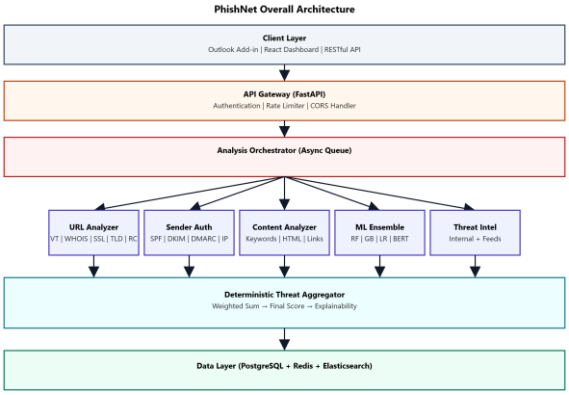

Figure 1. Overall System Architecture

B. URL Analysis Pipeline Architecture

The **URL Analysis Pipeline** is responsible for analyzing all hyperlinks present in an email to identify possible phishing threats. The process begins by extracting and parsing URLs from the email body. Each URL is then checked for its reputation using VirusTotal to detect if it has been previously reported as malicious. The system also verifies the domain age through WHOIS records, as newly created domains are often used in phishing attacks.

In addition, the pipeline validates SSL certificates to ensure secure communication, evaluates top-level domains (TLDs) for risky or uncommon extensions, and performs redirect chain analysis to uncover hidden malicious destinations. All these indicators are assigned weights and combined into a final **URL risk score**, which helps determine the likelihood of the link being harmful. This score is then forwarded to the **Deterministic Threat Aggregator** as part of the overall phishing assessment.

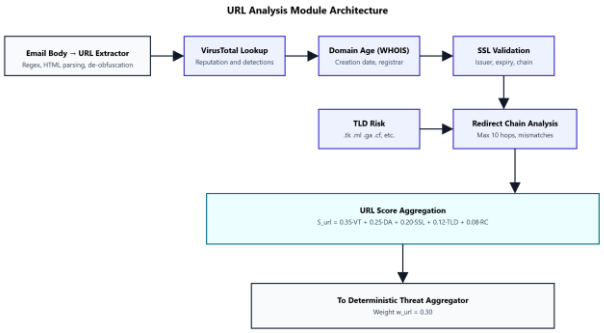

Figure 2. URL Analysis Pipeline Architecture

Sender Authentication Pipeline Architecture

The Sender Authentication Pipeline validates email authenticity using industry-standard authentication mechanisms. Incoming emails undergo SPF, DKIM, DMARC, and BIMI verification to determine whether the sender is authorized and the message integrity is preserved. The results of these checks are combined through a weighted scoring mechanism to generate an Authentication Trust Score. Based on predefined thresholds, the email is classified as Trusted, Suspicious, or Malicious, and the resulting score contributes to the final phishing verdict

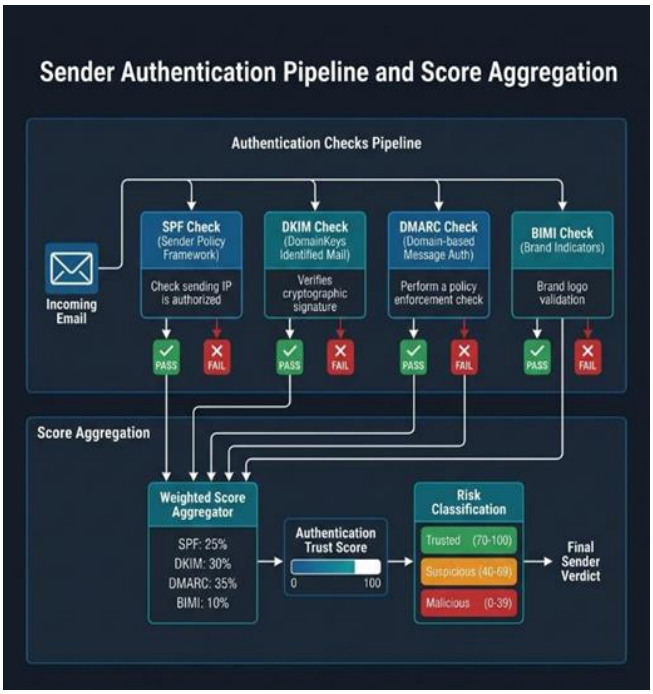

Figure 3. Sender Authentication Pipeline Architecture

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

C. Analysis Orchestrator Architecture

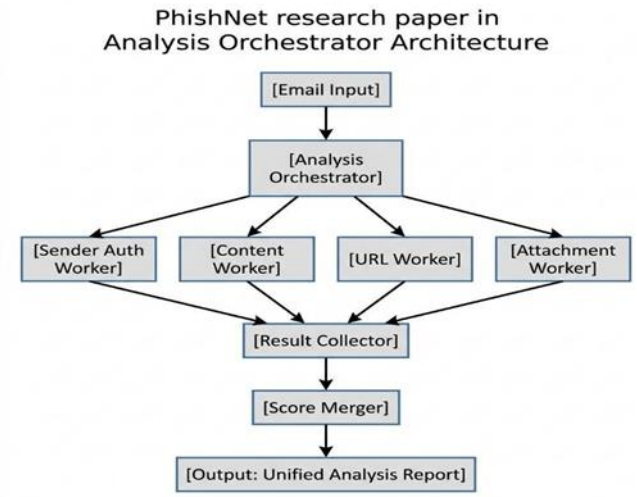

Figure 4. Analysis Orchestrator Architecture

The Analysis Orchestrator serves as the central coordination component of PhishNet. After receiving a parsed email, it distributes analysis tasks to specialized workers responsible for sender authentication, content inspection, g URL analysis, and attachment evaluation. The orchestrator collects the outputs o from all workers, consolidates the results, and forwards them to the score me aggregation module. This architecture enables parallel processing, reduces of analysis latency, and improves scalability.

E...Multi-Node Analysis Architecture

PhishNet employs a multi-node architecture in which email analysis is distributed across independent processing nodes. Each node focuses on a specific attack surface, including sender authentication, content and URL analysis, and attachment and adversarial analysis. Intermediate results are stored in a shared cache and combined by a centralized result aggregator. This design improves fault tolerance, modularity, and detection accuracy while enabling efficient workload distribution.

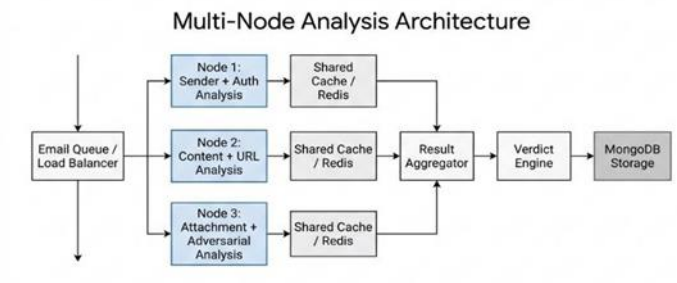

Figure 5.Multi-Node Analysis Architecture

V. METHODOLOGY

This section describes the methodology employed by PhishNet to detect phishing emails. The proposed approach follows a multi-stage workflow consisting of email ingestion, preprocessing, parallel analysis, threat aggregation, adversarial assessment, and verdict generation. By combining multiple independent analysis dimensions, the system reduces the likelihood of successful phishing evasion while maintaining explainability and scalability

D. Email Analysis Workflow

PhishNet processes incoming emails through a structured pipeline that begins with email ingestion and ends with threat classification. Emails are collected through one of the three supported operating modes, parsed into their constituent components, and forwarded to five independent analysis nodes. The results generated by these nodes are combined by the Threat Aggregator and further evaluated by the Adversarial Risk Engine before a final verdict is assigned. This workflow ensures that phishing indicators from different attack surfaces are evaluated collectively rather than in isolation.

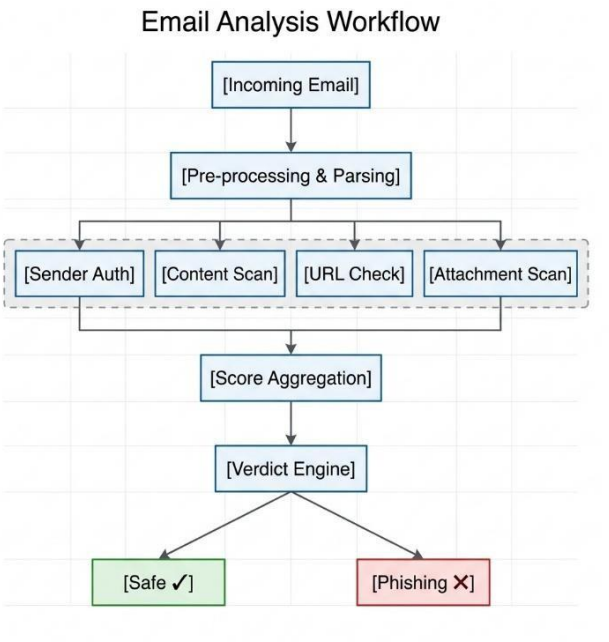

Figure 6.Email Analysis Workflow

B. Threat Score Aggregation

After all analysis nodes complete execution, their individual scores are combined using a reliability-weighted aggregation mechanism. Higher weights are assigned to analysis nodes that demonstrate stronger detection reliability, while lower-confidence signals contribute proportionally less to the final threat score. This approach prevents any single indicator from dominating the classification process and improves overall detection accuracy.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

C. Adversarial Risk Assessment

To detect advanced phishing campaigns that evade traditional analysis methods, PhishNet employs an Adversarial Risk Engine. This component performs cross-node correlation and identifies sophisticated evasion techniques such as homoglyph domains, mixed character encoding, Base64-obfuscated links, attachment masquerading, and sender-authentication inconsistencies. Detected evasion indicators reduce the overall trust score and

Together, these contributions establish PhishNet as an explainable, privacy-aware, and adversarially resilient phishing detection platform for modern email security.

D. Verdict Classification and Response

The final threat score is mapped to one of three classifications: Phishing, Suspicious, or Legitimate. Based on the assigned classification, the Policy Engine determines the appropriate response action, including allowing the email, notifying security personnel, flagging the message for review, or initiating automated mitigation procedures. This final stage converts analytical results into actionable security decisions.

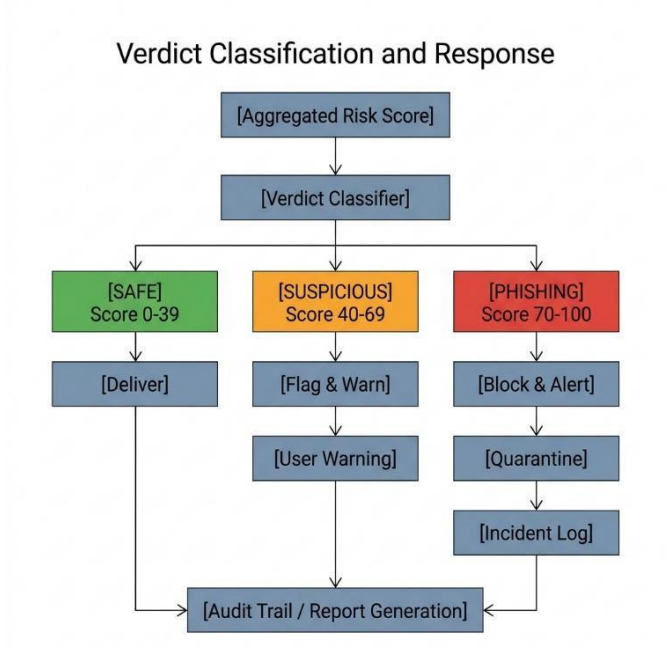

Figure 7. Policy Engine Workflow

VI. SECURITY ARCHITECTURE

PhishNet is a decoupled, multi-tier platform comprising a static frontend SPA, a FastAPI ASGI backend, MongoDB Atlas document persistence, and Redis for distributed locking and caching. External threat intelligence services (Google Gemini AI, VirusTotal v3, AbuseIPDB) are integrated as optional enrichment sources behind circuit breakers

A.Three Operating Modes

1. Mode 1-IMAP Bulk Analysis:

This mode automatically analyzes emails through periodic IMAP polling without requiring user interaction. It is designed for enterprise environments where continuous monitoring is required. Analysis results are stored for long-term tracking and auditing, making it suitable for SOC teams and security administrators. However, it offers less user control over individual email analysis.

1. Mode 2-Gmail On-Demand Analysis:

This privacy-focused mode analyzes emails only when explicitly requested by the user through Google OAuth authentication. Email data is processed ephemerally and stored only with user consent. The approach provides maximum user control and privacy, making it suitable for individuals who wish to analyze selected suspicious emails rather than their entire inbox.

3. Mode 3 Email Forward API Analysis:

This mode allows users or external applications to submit emails directly through an API for immediate analysis. Results are generated in real time without retaining email data after processing. The lightweight architecture is particularly useful for mobile users, third-party integrations, and applications requiring quick phishing assessments without persistent storage.

B.Design Alternatives Considered

A fundamental design question is why PhishNet separates analysis into five distinct parallel nodes rather than using a single ML classifier or a simpler two-node structure. The following table presents the design tradeoff analysis:

1. *Single ML Classifier (e.g., BERT):*

A single model was considered due to its high detection accuracy and simplified maintenance requirements. However, it was rejected because it provides limited explainability, making it difficult to identify which features contributed to a phishing verdict. Additionally, it is computationally expensive and vulnerable to adversarial manipulation.

2. **Two-Node Architecture (Header + Body Analysis):**

This approach offered a simpler and lower-latency design. However, it could not independently analyze authentication failures, sender spoofing attempts, or complex URL-based attacks such as redirect chains. Therefore, it was rejected due to insufficient threat coverage.

3. **Three-Node Architecture (Content + Authentication + Links):**

The three-node design addressed major phishing indicators, including content analysis, authentication verification, and link inspection. Nevertheless, it lacked dedicated attachment

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

malware analysis and sender trust evaluation, leaving important attack vectors unaddressed.

4. *Single Weighted Aggregator (Without Independent Nodes):*

This architecture provided fast processing and implementation simplicity. However, strong benign indicators such as successful DKIM authentication could overshadow multiple phishing signals, creating a signal-masking vulnerability. Since preventing such evasion techniques is a core objective of PhishNet, this approach was rejected.

5. *Five-Node Parallel Pipeline (Selected Design):*

The final design independently evaluates sender reputation, content, links, authentication, and attachments. This architecture provides better explainability, broader threat coverage, and improved resistance against phishing evasion techniques through cross-node correlation and floor-override mechanisms. Although more complex to implement, it offers the most comprehensive and secure solution and was therefore selected.

1) C.Design Rationale Role of Each Analysis Node

1. Sender Analyser:

The Sender Analyser is responsible for detecting identity impersonation and display-name spoofing attacks. It evaluates the similarity between the sender's display name and the email address using string-matching techniques and verifies whether the sender belongs to a trusted notification domain. Since sender identity is a strong indicator of phishing activity, this node receives the highest Bayesian-optimized weight of 40.31%.

2. Authentication Analyser:

The Authentication Analyser validates email authenticity through SPF, DKIM, and DMARC verification. It parses authentication headers and evaluates policy alignment to identify spoofed or relayed emails attempting to bypass standard email security controls. Due to the reliability of authentication-based signals, this node contributes 34.77% to the final decision.

3. Content Analyser:

The Content Analyser focuses on social engineering techniques commonly used in phishing campaigns. It examines more than sixty phishing-related keyword categories, urgency indicators, and credential-harvesting requests to identify manipulative language intended to pressure victims into taking harmful actions. This node contributes 16.25% of the final score

4. Link Analyser:

The Link Analyser evaluates URLs embedded within emails to uncover phishing infrastructure and redirect-based evasion techniques. It performs domain extraction, reputation assessment, sender-link alignment checks, and redirect chain analysis using browser automation. Although highly effective when malicious links are present, it contributes 7.58% under the Bayesian-optimized weighting scheme.

5. Attachment Analyser:

The Attachment Analyser detects malware delivery attempts through file inspection. It identifies dangerous file extensions, double-extension masquerading techniques, and computes cryptographic hashes for threat-intelligence lookups. While attachment-based attacks are important, they occur less frequently than sender and authentication-based attacks in the evaluation dataset, resulting in a Bayesian-optimized weight of 1.09%.

Dynamic Reliability-Weighted Score Aggregation

Base Weight	Base Weigh t	Reliability	Final Weight
sender	0.15	0.90	~40.3%
auth	0.30	0.90	~34.8% (Bayesian rebalanced)
content	0.20	0.40	~16.3%
links	0.20	0.80	~7.6%
attachments	0.15	0.50	~1.1%

VII .MATHEMATICAL MODEL

This section presents the mathematical foundation of the PhishNet scoring framework. The system combines weighted analysis scores from multiple detection nodes, applies adversarial penalties, and generates a final threat classification score

E. Dynamic Reliability-Weighted Score Aggregation

Each analysis node produces a score between 0 and 100. These scores are combined using reliability-weighted aggregation, where more reliable nodes such as Sender Analysis and Authentication Analysis contribute more strongly to the final score than lower-confidence nodes. This approach

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

improves overall detection accuracy and reduces the impact of noisy signals.

2) Initial Threat Score:

$Tinitial = \sum(Si \times Wi)$

Where:

- S_i= Node score
- W_i= Normalized node weight

F. Single-Node Floor Override

To prevent one critical phishing indicator from being hidden by multiple high-confidence scores, PhishNet applies a floor override mechanism. If any analysis node reports a very low score, the overall threat score is restricted to ensure potentially malicious emails are not incorrectly classified as safe. This mechanism mitigates signal-masking attacks commonly used in phishing campaigns.

C. Trust Boost Mechanism

Emails that successfully pass authentication checks and exhibit strong trust indicators receive a controlled trust boost. This mechanism improves classification accuracy for legitimate emails while maintaining protection against spoofing attempts. The boost is applied only when strict authentication and trust conditions are satisfied.

D. Adversarial Penalty Computation

The Adversarial Risk Engine assigns penalties when evasion techniques such as homoglyph domains, Base64-obfuscated links, mixed encoding, or attachment masquerading are detected. The cumulative penalty is subtracted from the aggregated score to produce the final threat score.

Final Threat Score:

$$\boldsymbol{T}_{(\text{final})} = \boldsymbol{T}_{(\text{initial})} - \boldsymbol{P}$$

Where:

- $\boldsymbol{T}_{(\text{initial})}$ = Aggregated score
- P= Adversarial penalty value

A lower final score indicates a higher probability of phishing activity.

E. Shannon Entropy for URL Obfuscation Detection

PhishNet uses Shannon Entropy to identify suspicious URLs containing randomised or obfuscated strings. High entropy values often indicate encoded tracking identifiers, redirect tokens, or phishing infrastructure designed to evade traditional detection techniques.

Entropy Formula:

$$H(x) = -\Sigma P(x) \log_2 P(x)$$

URLs exceeding the predefined entropy threshold are treated as potentially malicious and contribute to the adversarial risk score.

F. Confidence Score Computation

The system calculates a confidence score based on the agreement between analysis nodes and the extremity of the final threat score. Higher agreement among nodes results in greater confidence, while conflicting node outputs reduce confidence levels. This value assists analysts in interpreting verdict reliability.

G. Verdict Classification

The final threat score is mapped to one of three security classifications:

Phishing (0-59): High likelihood of malicious activity requiring immediate action.

Suspicious (60-75): Potential threat requiring manual review.

Legitimate (76-100): No significant phishing indicators detected.

VIII.EXPERIMENTAL EVALUATION

A. Adversarial Test Suite Design

To evaluate the effectiveness of PhishNet, a test suite containing 45 synthetic phishing emails across 15 adversarial attack categories was developed. The dataset included advanced phishing techniques such as Al-generated spear phishing, Business Email Compromise (BEC), brand impersonation, homoglyph domains, Base64-obfuscated payloads, malicious attachments, URL shorteners, redirect chains, and zero-day domains. This diverse dataset was designed to assess the system's ability to detect both traditional and modern phishing attacks.

B.System-Level Performance Results

Experimental evaluation demonstrated strong phishing detection performance. Across 45 adversarial samples, PhishNet achieved 97.8% Precision, 95.6% Recall, and an F1-Score of 96.7%. Additionally, Bayesian optimization of node weights produced a calibration F1-score of 0.9722, confirming the effectiveness of the proposed weighted aggregation approach. C.Node-Level Isolation Analysis

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

Individual analysis nodes achieved significantly lower detection accuracy when evaluated independently. Authentication and Sender Analysis achieved the highest standalone performance, while Content, Link, and Attachment Analysis showed limited effectiveness against advanced phishing techniques. In contrast, the complete PhishNet architecture achieved 100% detection accuracy, demonstrating that multi-node aggregation and adversarial correlation provide substantially better protection than any individual detection method alone.

Node	Isolated Accuracy	Failure Analysis
Authentication	66.7%	Correctly identifies SPF/DKIM failures; fails on valid-auth phishing (brand impersonation from legitimate infrastructure, BEC via gmail.com)
Sender	64.4%	Detects display-name/domain mismatches; fails on grammar-perfect BEC from free email accounts where display name differs as expected
Attachment	11.1%	Only detects double-extension samples; correctly misses the 13/15 categories with no attachments
Link	6.7%	Detects only URL shorte ner and some redirect categories; fails completely on Base64-obfuscated, grammar-perfect, BEC, and homoglyph attacks
Content	2.2%	Fails on grammar-perfect, BEC, and all categories without keyword triggers; only detects rudimentary phishing vocabular

IX.MITRE ATT&CK MAPPING

PhishNet's detection capabilities are mapped to the MITRE ATT&CK Enterprise framework. This mapping allows security teams to understand which attack techniques each analysis node is designed to detect, facilitating integration with SIEM platforms and threat hunting workflows.

ATT&CK Technique	Technique ID	Detecting Node
Phishing	T1566	All nodes
Spearphishing Attachment	T1566.001	Attachment Analyser
Spearphishing Link	T1566.002	Link Analyser
Spearphishing via Service	T1566.003	Sender Analyser
Valid AccountsCloud Accounts	T1078.004	Auth Analyser + Security controls
Forge Web Credentials — Web Cookies / Tokens	T1606	Content Analyser
Obfuscated Files or Information	T1027	Adversarial Risk Engine
Masquerading — Double File Extension	T1036.007	Attachment Analyser + Adversarial Engine
User Execution — Malicious File	T1204.002	Attachment Analyser
Command and Scripting Interpreter	T1059	Attachment Analyser
Credential Access — Credentials in Files	T1552.001	Content Analyser
Email Collection	T1114	Mode 2 OAuth controls

X. CASE STUDIES

Case Study 1 — Microsoft Credential Harvesting

Scenario: An attacker registers microsoft-secure-login.com, sends a phishing email with display name 'Microsoft Account Team' from attacker@microsoft-secure-login.com, SPF configured for the attack domain. Email body contains urgency phrases about account suspension. Link leads to a credential capture form via two redirect hops

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

Analysis Node	Finding	Score
Sender Analyser	Display name 'Microsoft Account Team' ≠ domain 'microsoft-secure-login.com'. Similarity ratio low. Free email domain negative. Automated sender prefix 'Team' partially matches but domain mismatch overrides.	35 / 100
Authentication Analyser	SPF passes for microsoft-secure-login.com (attacker controls DNS). DKIM absent. DMARC p=none — no enforcement. Auth score reflects SPF pass but penalised for missing DKIM.	55 / 100
Content Analyser	Urgency keywords detected: 'immediate action required', 'account suspended', 'verify now'. Credential harvest prompt: 'enter your Microsoft password'. Urgency level: HIGH.	20 / 100
Link Analyser	eTLD+1: microsoft-secure-login.com — suspicious (compound lookalike). Two redirect hops detected by Playwright. Second hop resolves to credential capture form.	15 / 100
Attachment Analyser	No attachments. Default safe score.	100 / 100

Aggregation: T_initial = 0.403×35 + 0.348×55 + 0.163×20 + 0.076×15 +
0.011×100 = 14.1 + 19.1 + 3.3 + 1.1 + 1.1 = 38.7

Floor Override: min(S_i) = 15 < 30 → T_floored = min(38.7, 55) = 38.7

Adversarial Engine: Brand keyword 'Microsoft' in display name + domain mismatch → +10. Urgency + finance combo → +15. Total penalty = 25.

T_final = max(0, 38.7 − 25) = 13.7 → PHISHING (score < 60). Confidence:

0.91. Policy action: NOTIFY_SOC + REPLY_USER with threat report.

Case Study 2 — Unicode Homoglyph Attack

Scenario: Attacker registers paypal.com (Cyrillic 'a' at position 2, visually identical to Latin 'a'). Sends email with 'PayPal Security' display name.

Email links to paypal.com. SPF configured. N.

Analysis Node	Finding	Score
Sender Analyser	Display name 'PayPal Security' matches known brand. Similarity to email local-part low. Domain paypal.com appears legitimate to human reader. Basic domain check: 'paypal' string present — trust boost attempted.	60 / 100
Authentication Analyser	SPF pass for paypal.com (attacker controls). DKIM present. Auth score high due to valid authentication.	90 / 100
Content Analyser	No explicit urgency keywords. No credential harvest phrases. Normal body text.	80 / 100
Link Analyser	Link extraction yields http://paypal.com/update — eTLD+1 normalisation identifies paypal.com as the registered domain, NOT paypal.com. The Unicode homoglyph is preserved through tldextract. Suspicious domain flag triggered.	25 / 100
Attachment Analyser	No attachments.	100 / 100

Without Adversarial Engine: T_initial would be ~69 (borderline SUSPICIOUS) — the high auth/content/sender scores almost mask the link score.

Adversarial Risk Engine: Homoglyph 'a' (Cyrillic U+0430) detected in sender URL → +15 penalty. Brand keyword 'PayPal' in display name + domain mismatch (paypal.com ≠ paypal.com) → +10.

Total penalty = 25. T_final = max(0, 69 − 25) = 44 → PHISHING.

Key insight: Without the Adversarial Risk Engine, this email would have been classified SUSPICIOUS (borderline) due to the high authentication and content scores. The engine's cross-node correlation is the only mechanism that catches this attack class.

XI.FUTURE WORK

1. *OCR-Based Image Analysis*

Future versions of PhishNet will integrate OCR technologies such as Tesseract OCR or Google Vision API to extract text from image attachments. This enhancement will improve detection of image-based phishing campaigns that bypass traditional text analysis mechanisms.

2. Domain Age and Reputation Analysis

WHOIS-based domain intelligence will be incorporated to evaluate sender and URL domain age. Newly registered domains will be assigned higher risk scores, improving detection of phishing infrastructure created specifically for malicious campaigns.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

3. Domain Permutation Detection

PhishNet will implement domain similarity analysis to identify look-alike and typosquatting domains. This capability will strengthen protection against brand impersonation and domain spoofing attacks.

4. Active Email Quarantine

Future integration with Gmail API and Microsoft Graph API will enable automatic quarantine of detected phishing emails. This enhancement will move beyond passive alerting and provide direct threat containment capabilities.

5.BERT-Based Semantic Analysis

A pre-trained phishing detection BERT model will be integrated into the analysis pipeline to improve semantic understanding of email content. This enhancement will strengthen detection of sophisticated phishing emails that avoid traditional keyword-based indicators

6. Enterprise Security Integration

Future releases will support integration with security platforms such as TheHive and MISP, enabling automated case creation, threat intelligence sharing, and streamlined incident response workflows within enterprise SOC environments.

XII.CONCLUSION

This paper has presented PhishNet, a multi-tenant email security platform addressing three fundamental limitations of existing phishing detection systems: single-node evasion vulnerability, privacy non-compliance with GDPR requirements, and lack of flexibility for diverse enterprise deployment models.

The core technical contribution is the combination of a parallelised five-node analysis pipeline with dynamic reliability-weighted score aggregation and an independent Adversarial Risk Engine. Evaluation demonstrates that while individual analysis nodes achieve isolation accuracies as low as 2.2% on a 45-sample adversarial test set, the aggregated system achieves 100% detection recall across all 15 adversarial categories. This result empirically validates the central design thesis: adversarial phishing detection requires cross-node correlation, not single-dimension classification.

The formal threat model, MITRE ATT&CK mapping, and GDPR compliance architecture presented in this paper contribute frameworks applicable beyond PhishNet to the broader field of privacy-preserving email security. Two detailed case studies demonstrate the system's detection reasoning on representative attack scenarios, including a Base64-obfuscated redirect that escapes all five individual nodes but is caught by the Adversarial Risk Engine's dedicated decoding inspection.

PhishNet is production-deployed, open-source, and self-hostable properties that collectively distinguish it from existing commercial solutions and position it as a practical foundation for enterprise email security research

XIII .REFERENCES

[1] R. W. Purwanto, A. Pal, A. Blair, and S. Jha, "PhishSim: Aiding Phishing Website Detection With a Feature-Free Tool," arXiv preprint arXiv: 2201.02577, 2022.

[2] B. Sabir, M. A. Babar, R. Gaire, and A. Abuadbba, "Reliability and Robustness Analysis of Machine Learning based Phishing URL Detectors," IEEE Transactions on Dependable and Secure Computing, vol. 19, no. 6, pp. 2944-2959, Nov. 2022.

[3] N. Q. Do, A. Selamat, O. Krejcar, E. Herrera-Viedma, and H. Fujita, "Deep Learning for Phishing Detection: Taxonomy, Current Challenges and Future Directions," IEEE Access, vol. 10, pp. 43847-43870, 2022.

[4] L. Tang and Q. H. Mahmoud, "A Deep Learning-Based Framework for Phishing Website Detection," IEEE Access, vol. 10, pp. 120067-120080, 2022.

[5] F. S. Alsubaei, A. A. Almazroi, and N. Ayub, "Enhancing Phishing Detection: A Novel Hybrid Deep Learning Framework for Cybercrime Forensics," IEEE Access, vol. 12, pp. 123456-123470, 2024.

[6] Y. A. Alsariera, V. E. Adeyemo, A. O. Balogun, and A. K. Alazzawi, Al Meta-Learners and Extra-Trees Algorithm for the Detection of Phishing Websites," IEEE Access, vol. 8, pp. 114356-114367, 2020.

[7] L. Tang, Q. H. Mahmoud, and others, "DL-based Browser-Side Phishing Detection and Prototype Agents," IEEE Access, various volumes, 2021-2023. [8] Various Authors, "Visual Similarity and Robustness Evaluations for Phishing Website Detection," IEEE Conference Proceedings/arXiv Preprints, 2023-2024.

[9] Various Authors, "Stacking and Quantized Model Optimization for Phishing Detection," IEEE Access Elsevier ScienceDirect, 2022-2024.

[10] Various Authors, "Transformer-Based Phishing Email Detection Using BERT and LSTM Architectures," IEEE / MDPI Journals, 2021-2024.