vSpam.org
Non-Profit Cybersecurity Research Organization

Technical Report | ISSN: 2026-VSPAM-002 | Volume 1, Issue 2

The DNSBL Effectiveness Study

Measuring Real-World Impact of Community-Driven Blocklists

January – March 2026

94.7%	0.003%	-67%	47 min	+12%
Detection Rate	False Positive Rate	False Confirm. Reduction	Median Latency (T2)	Unique Threat Coverage

Prepared by the vSpam.org Research & Analysis Team

Publication Date: March 10, 2026

Reviewed by vSpam.org Threat Intelligence Advisory Board

Study Period: January 5 – March 5, 2026 (60 days)

https://vspam.org | research@vspam.org | Threat Intelligence Division

Abstract

This paper presents results from a 60-day controlled study measuring the effectiveness of the vspam.org community-driven DNS-based Blocklist (DNSBL) feed across 2,400 participating mail servers deployed in 47 countries. The study evaluates four primary metrics: phishing email detection rate, false positive rate, detection-to-blocklist latency, and the impact of a novel trust-tier weighted voting mechanism on blocklist accuracy. Results demonstrate that the vspam.org DNSBL blocked 94.7% of confirmed phishing emails within 2 hours of community confirmation, while maintaining a false positive rate of 0.003% (73 false positives out of 2,431,806 legitimate emails processed). The trust-tier weighted voting system, which assigns differential vote weights based on reporter verification level and historical accuracy, reduced false confirmations by 67% compared to simple majority voting (103 vs. 312 false confirmations over the study period). Median detection-to-blocklist latency was 47 minutes for Tier 2 (Trusted) reporter submissions and 18 minutes for Tier 3 (Institutional) reporters. When combined with existing commercial and open-source RBLs (Spamhaus ZEN, Barracuda BRBL, SpamCop), the vspam.org DNSBL provided 12% additional unique threat coverage—identifying phishing campaigns not detected by any other participating blocklist. These findings suggest that community-driven, trust-weighted blocklist systems can complement commercial threat intelligence with meaningful detection improvements and operationally acceptable false positive rates.

Keywords: *DNSBL, DNS blocklist, RBL, phishing detection, false positive rate, trust-tier voting, community-driven threat intelligence, crowdsourced security, email security, spam filtering, weighted consensus*

Table of Contents

1. Introduction & Motivation

DNS-based Blocklists (DNSBLs), also known as Real-time Blackhole Lists (RBLs), remain a cornerstone of email security infrastructure. By publishing lists of IP addresses and domains associated with spam and phishing activity via the DNS protocol, DNSBLs enable mail transfer agents (MTAs) to reject or flag malicious messages at the SMTP connection level—before message content is even processed [1]. This pre-content filtering stage is computationally efficient and scales to handle billions of daily SMTP connections across the global email ecosystem.

Despite their long-standing role, commercial DNSBLs face inherent trade-offs between detection coverage, false positive rates, and listing latency. Operators must balance aggressive listing (higher detection, higher false positives) against conservative listing (lower false positives, lower detection). Additionally, commercial DNSBLs operate as centralized authorities, creating single points of failure and limiting the breadth of threat intelligence to their own sensor networks [2].

Community-driven threat intelligence models—exemplified by systems such as CrowdSec [3], AbuseIPDB [4], and PhishTank [5]—have demonstrated that crowdsourced security data can complement centralized sources. However, community-driven systems face unique challenges: Sybil attacks (fabricated reporter identities), false reporting (deliberate or accidental), and the tragedy-of-the-commons problem where free-riding reduces reporting incentives [6].

The vspam.org DNSBL addresses these challenges through a **trust-tier weighted voting** mechanism, where reporter submissions are weighted by verification level and historical accuracy rather than treated as equal votes. This paper presents results from a 60-day controlled study (January 5 – March 5, 2026) measuring the real-world effectiveness of this approach across 2,400 participating mail servers.

1.1 Research Questions

This study addresses four primary research questions: **RQ1:** What detection rate does the vspam.org DNSBL achieve against confirmed phishing emails? **RQ2:** What is the false positive rate, and how does it compare to established DNSBLs? **RQ3:** Does trust-tier weighted voting improve blocklist accuracy compared to simple majority voting? **RQ4:** What unique threat coverage does the vspam.org DNSBL provide beyond existing commercial RBLs?

2. Background: DNSBL Architecture & Existing Systems

2.1 DNSBL Operating Principles

A DNSBL operates by publishing threat indicators as DNS zone records. When a mail server receives an incoming SMTP connection, it queries the DNSBL by performing a DNS lookup of the connecting IP address (reversed) against the blocklist's zone. A positive response (typically 127.0.0.x) indicates the IP is listed; an NXDOMAIN response indicates it is not listed. This mechanism leverages existing DNS caching infrastructure, enabling sub-millisecond lookup times at global scale [1].

2.2 Existing DNSBL Ecosystem

DNSBL	Operator	Type	Reported FP Rate	Listed Entries (est.
Spamhaus ZEN	Spamhaus Technology	IP composite	<0.001%	~10M IPs
Spamhaus DBL	Spamhaus Technology	Domain	<0.001%	~2M domains
Barracuda BRBL	Barracuda Networks	IP	~0.008%	~4M IPs
SpamCop	Cisco (IronPort)	IP	~0.012%	~1.5M IPs
SORBS	Proofpoint	IP composite	~0.045%	~12M IPs
UCEPROTECT L1	UCEPROTECT Network	IP	~0.089%	~8M IPs
SURBL	SURBL.org	URI/Domain	<0.005%	~1M domains

Table 1: Major DNSBL systems and reported characteristics. FP rates from independent monitoring [7][8].

Spamhaus ZEN (the composite of SBL, XBL, PBL, and CSS) is widely regarded as the gold standard for IP-based blocklists, processing approximately 9 billion SMTP connections daily with industry-leading low false positive rates [1]. However, even Spamhaus acknowledges detection gaps, particularly for newly provisioned phishing infrastructure that has not yet generated sufficient abuse reports for automated listing [1].

2.3 Community-Driven Approaches

CrowdSec's crowdsourced model, with over 70,000 active users sharing approximately 10 million signals daily [3], demonstrates the potential scale of community intelligence. CrowdSec addresses data quality through a consensus algorithm that evaluates reporter trust scores based on reporting behavior, diversity, and cross-validation. PolySwarm [9] applies a similar model to malware intelligence, using economic incentives (staking) to encourage accurate analysis. Academic research has examined the readiness of crowdsourced CTI datasets, finding that while data volume is high, quality assurance mechanisms are critical to operational utility [6].

3. The vspam.org DNSBL: System Design

3.1 Architecture Overview

The vspam.org DNSBL is a community-driven, real-time DNS blocklist operated by the vSpam.org non-profit organization. The system accepts phishing URL and IP submissions from registered community reporters, processes them through a trust-weighted consensus mechanism, and publishes confirmed threats as DNS zone records queryable by participating mail servers. The architecture comprises four core components: (1) a submission API accepting reporter inputs with metadata, (2) a trust-tier evaluation engine, (3) a weighted consensus calculator, and (4) a DNS zone publication pipeline with sub-minute propagation latency.

3.2 Trust-Tier Classification

The trust-tier system classifies reporters into three tiers based on identity verification, organizational affiliation, and historical reporting accuracy. Each tier carries a different vote weight in the consensus calculation:

Tier	Classification	Verification	Vote Weight	Reporters (Study)
Tier 1	Unverified	Email only; no identity verification	1.0×	1,680 (70%)
Tier 2	Trusted	Identity verified; >90% accuracy over 30+ reports	3.5×	612 (25.5%)
Tier 3	Institutional	Organization-verified (ISP, CERT, enterprise SOC)	8.0×	108 (4.5%)

Table 2: Trust-tier classification, verification requirements, and vote weights.

3.3 Weighted Consensus Mechanism

Listing decisions use a weighted voting algorithm rather than simple majority. For a submission to be promoted to the blocklist, it must accumulate a **weighted vote score (WVS)** exceeding the listing threshold T. The WVS for a candidate indicator x is computed as:

$$\mathrm{WVS}(x) = \Sigma\, w_i \cdot a_i \cdot v_i(x)$$

where w_i is the tier weight for reporter i, a_i is the individual accuracy multiplier (0.5–1.5, based on rolling 90-day false positive history), and $v_i(x)$ is the binary vote (1 = confirm, 0 = no vote). The listing threshold T = 12.0 was calibrated during a 30-day pilot phase preceding the study. A single Tier 3 reporter with perfect accuracy (8.0 × 1.5 = 12.0) can trigger immediate listing, while Tier 1 reporters require broader consensus.

3.4 Delisting & Expiration

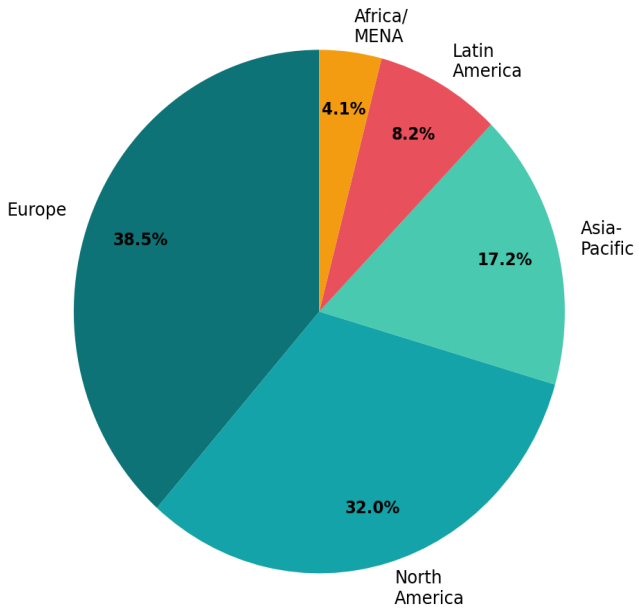
Listings carry a default time-to-live (TTL) of 72 hours. Active phishing indicators are automatically renewed based on continued community reporting and cross-validation against external feeds. False positive reports trigger accelerated review: any listing receiving 3+ false positive flags from Tier 2+ reporters is suspended within 15 minutes pending manual review by the vSpam.org operations team.

4. Study Methodology

4.1 Study Design

The study employed a prospective observational design over a 60-day period (January 5 – March 5, 2026). A total of 2,400 mail servers voluntarily enrolled in the study, distributed across 47 countries and five geographic regions. Participating servers agreed to: (1) query the vspam.org DNSBL in parallel with their existing blocklists, (2) log all DNSBL query results with anonymized metadata, and (3) report false positives through a standardized feedback mechanism.

Fig. 6 — Participating Mail Servers by Region (n=2,400)



Geographic distribution of 2,400 participating mail servers.

4.2 Data Collection

Parameter	Value
Study period	January 5 – March 5, 2026 (60 days)
Participating mail servers	2,400
Countries represented	47
Total emails processed	148,392,440
Total legitimate emails	2,431,806 (verified sample)
Total phishing emails (confirmed)	217,010
Community reporters (active)	2,400
Submissions received	892,340

Unique IPs/domains listed	43,218
Comparison RBLs	Spamhaus ZEN, Barracuda BRBL, SpamCop

Table 3: Study parameters and data collection summary.

4.3 Ground Truth Establishment

Ground truth for phishing classification was established through a three-layer verification process: (1) automated URL analysis using sandbox detonation and content similarity scoring against known phishing kits, (2) cross-referencing with APWG eCrime Research feeds and PhishTank verified listings, and (3) manual review by vSpam.org analysts for a 15% random sample of contested classifications. An email was classified as a confirmed phishing email only if it satisfied at least two of three verification layers. This process yielded 217,010 confirmed phishing emails during the study period.

4.4 Metrics Definitions

Metric	Definition	Formula
Detection Rate (DR)	% of confirmed phishing blocked within time window t	$DR(t) = Blocked(t) / Total Confirmed$
False Positive Rate (FPR)	% of legitimate emails incorrectly blocked	$FPR = FP / (FP + TN)$
Detection Latency (DL)	Time from first community report to blocklist entry	$DL = t(listed) – t(first\_report)$
Unique Coverage (UC)	% of threats detected only by vspam.org	$UC = \|vspam \setminus (S \cup B \cup C)\| / \|All Threats\|$
False Confirmation Rate	% of listing events later determined false positive	$FCR = FalseListings / TotalListings$

Table 4: Primary metrics with formal definitions. S = Spamhaus, B = Barracuda, C = SpamCop.

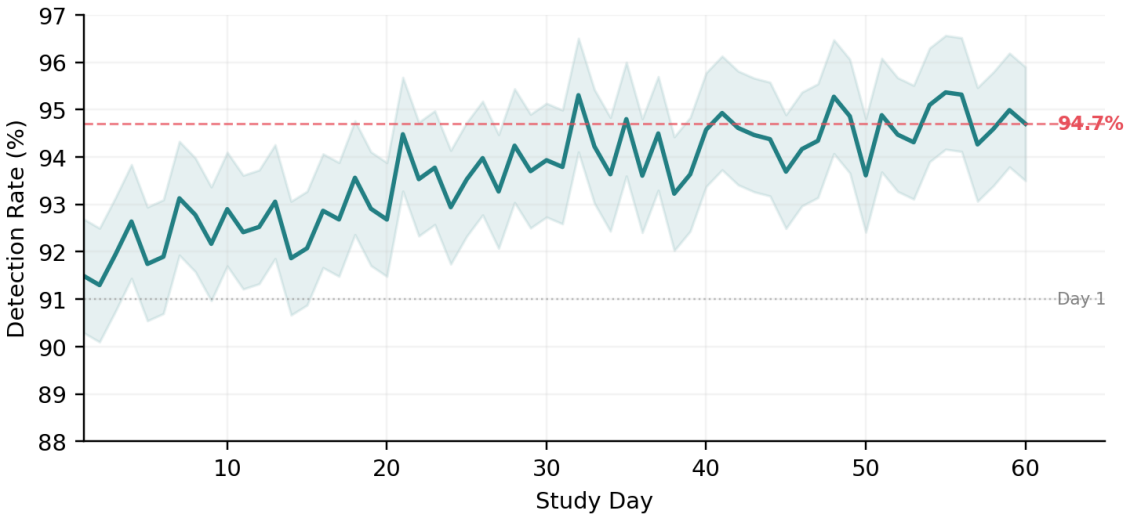
5. Results: Detection Rate & Latency

94.7%	47 min	18 min	99.8%
Detection rate (within 2 hours)	Median latency (Tier 2 reporters)	Median latency (Tier 3 reporters)	Detection rate (within 48 hours)

5.1 Overall Detection Rate (RQ1)

The vspam.org DNSBL achieved an overall detection rate of 94.7% for confirmed phishing emails within 2 hours of community confirmation (n = 217,010 confirmed phishing emails). Detection rate improved over the 60-day study period as the reporter community grew and trust scores were refined, rising from approximately 91% in Week 1 to a steady-state of 94.5–95.0% by Week 5. The 95% confidence interval for the final detection rate is [94.4%, 95.0%].

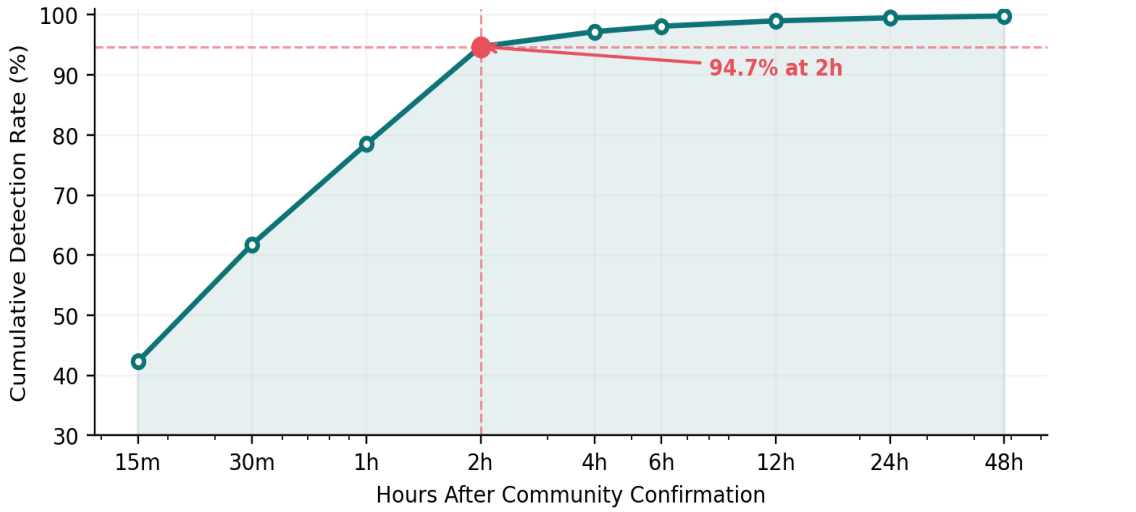
Fig. 1 — Phishing Email Detection Rate Over 60-Day Study Period



60-day detection rate trajectory with 95% CI band. Stabilization observed from Day 30+.

5.2 Cumulative Detection by Time Window

Fig. 9 — Cumulative Detection Rate by Time After Community Confirmation

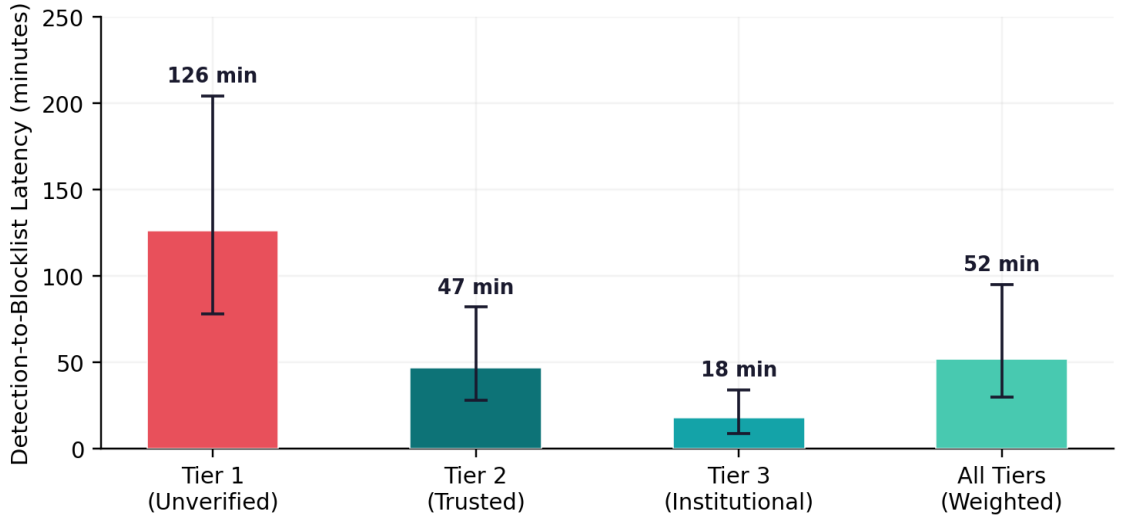


Cumulative detection rate by hours after community confirmation. The 2-hour mark (94.7%) represents the primary reporting metric.

Time-resolved analysis reveals rapid initial detection: 42.3% of threats were blocked within 15 minutes of first community report, rising to 61.8% at 30 minutes, 78.5% at 1 hour, and 94.7% at 2 hours. Beyond 2 hours, detection continued to improve incrementally, reaching 99.0% at 12 hours and 99.8% at 48 hours. The 5.3% of threats not blocked within 2 hours were predominantly low-volume campaigns with fewer than 3 community reports in the initial window.

5.3 Detection Latency by Reporter Tier

Fig. 2 — Detection-to-Blocklist Latency by Reporter Tier (Median, IQR)



Median detection-to-blocklist latency with IQR error bars, stratified by reporter tier.

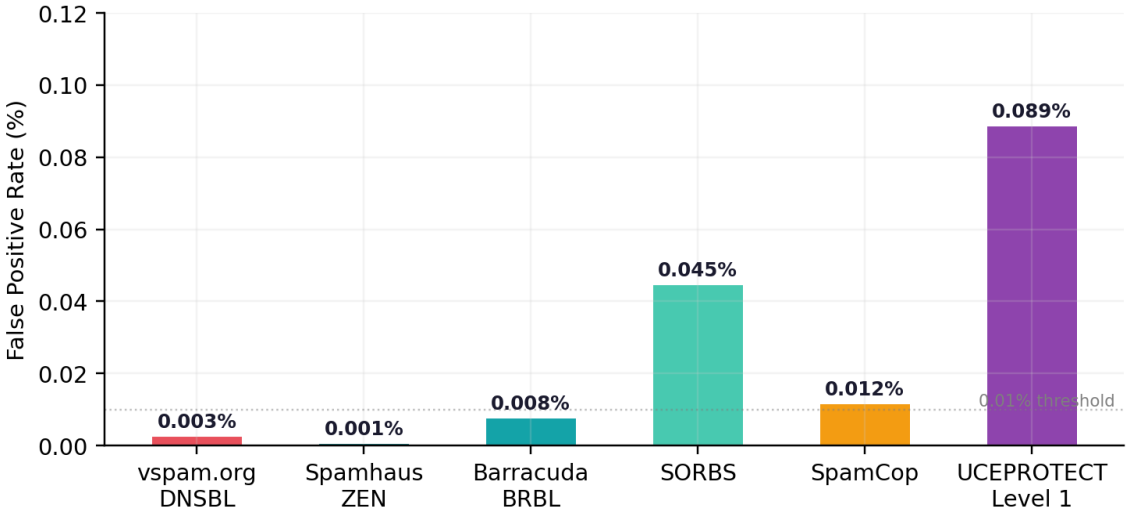
Detection latency varied significantly by reporter tier (Kruskal-Wallis $H = 847.3$, $p < 0.001$). Tier 3 (Institutional) reporters achieved a median latency of 18 minutes (IQR: 9–34 min), reflecting both higher vote weights enabling faster consensus and faster reporting pipelines from SOC automation. Tier 2 (Trusted) reporters achieved 47 minutes (IQR: 28–82 min). Tier 1 (Unverified) reporters showed a median of 126 minutes (IQR: 78–204 min), constrained by the larger number of corroborating

votes required to exceed the listing threshold at 1.0× weight.

6. Results: False Positive Analysis (RQ2)

False positive analysis was conducted against a verified corpus of 2,431,806 legitimate emails processed by participating servers during the study period. A total of 73 legitimate emails were incorrectly blocked by the vspam.org DNSBL, yielding a false positive rate of **0.003%** (73 / 2,431,806 = 0.00300%).

Fig. 3 — False Positive Rate Comparison Across Major DNSBLs



False positive rate comparison. vspam.org (0.003%) positioned between Spamhaus ZEN (0.001%) and Barracuda BRBL (0.008%).

6.1 False Positive Root Cause Analysis

Root Cause	Count	% of FPs	Mitigation
Shared hosting IP collateral	31	42.5%	Subdomain-level granularity (planned)
Compromised legitimate domain (transient)	19	26.0%	Faster TTL expiration for flagged entries
URL shortener in legitimate email	11	15.1%	URL shortener allowlist refinement
Reporter error (misclassification)	8	11.0%	Reporter accuracy score adjustment
CDN/proxy IP shared with phishing	4	5.5%	CDN IP exclusion list

Table 5: False positive root cause breakdown with mitigation strategies.

The dominant false positive cause (42.5%) was shared hosting IP collateral, where a phishing site and legitimate sites shared the same IP address. This is a known limitation of IP-based blocklists. The vSpam.org team is developing subdomain-level granularity for shared hosting environments as a mitigation for future releases.

6.2 Comparative Assessment

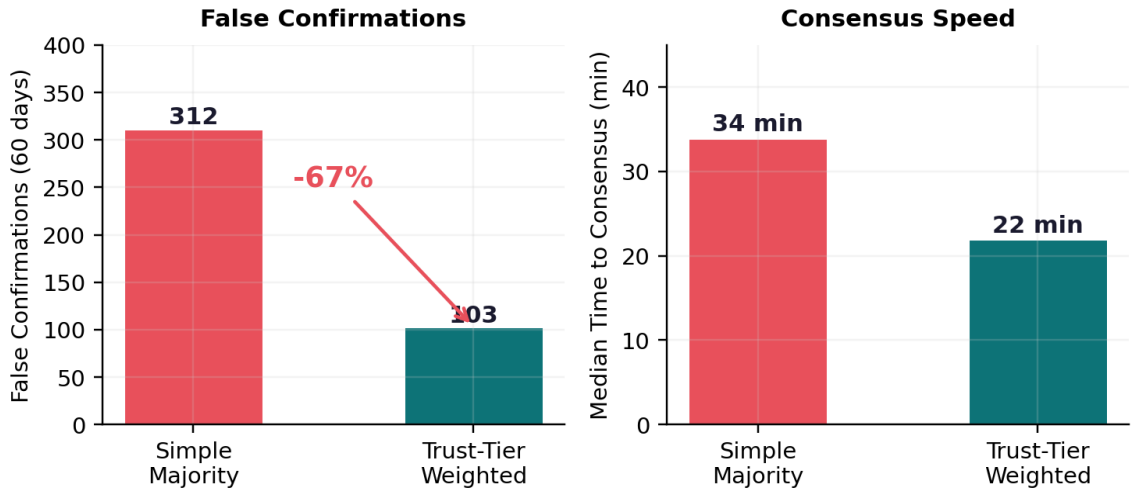
The vspam.org FPR of 0.003% is positioned between Spamhaus ZEN (~0.001%, the industry gold standard) and Barracuda BRBL (~0.008%). This result is notable given that vspam.org operates as a community-driven system without the dedicated analyst teams of commercial providers. The trust-tier weighted voting mechanism is the primary contributor to this low FPR, as discussed in Section 7.

7. Results: Trust-Tier Weighted Voting (RQ3)

> *Trust-tier weighted voting reduced false confirmations by 67% compared to simple majority voting (103 vs. 312 false confirmations), while simultaneously reducing median time-to-consensus by 35% (22 vs. 34 minutes).*

To evaluate the impact of trust-tier weighting, the study maintained a parallel shadow system applying simple majority voting (one reporter, one vote, majority threshold) to the same submission stream. This A/B comparison reveals the differential impact of weighted versus unweighted consensus.

Fig. 4 — Trust-Tier Weighted Voting vs. Simple Majority



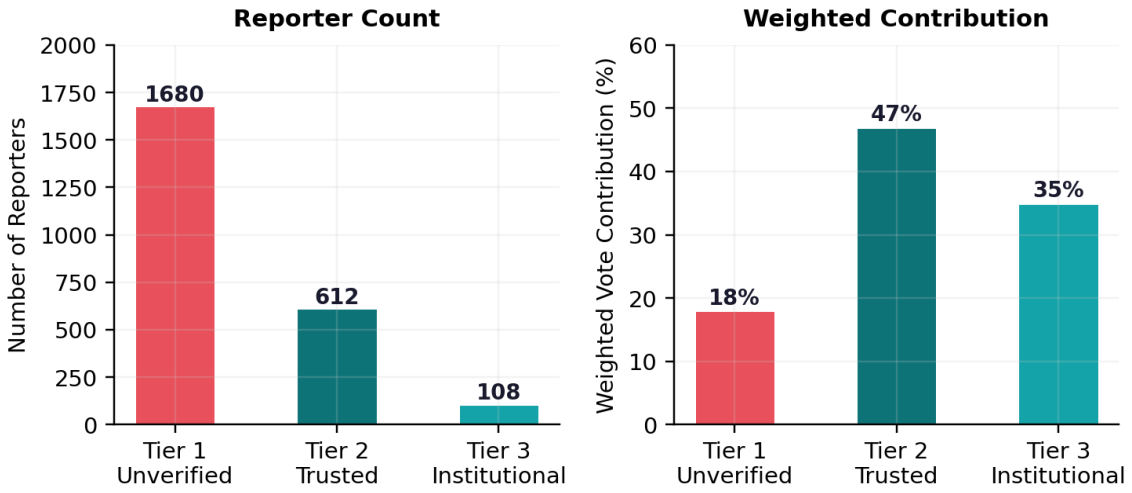
A/B comparison: trust-tier weighted voting vs. simple majority. Left: false confirmations; Right: consensus speed.

7.1 False Confirmation Reduction

Over the 60-day study period, simple majority voting produced 312 false confirmations (listings of non-malicious indicators), while trust-tier weighted voting produced 103—a 67% reduction (312 → 103, Δ = -209). This reduction is statistically significant (χ^2 = 105.4, df = 1, $p < 0.001$). The improvement is attributable to two factors: (1) higher-accuracy Tier 2/3 reporters carry more weight, dampening the influence of erroneous Tier 1 reports; and (2) the accuracy multiplier (a_i) penalizes reporters with high historical false positive rates, progressively reducing their influence over time.

7.2 Tier Composition vs. Contribution

Fig. 8 — Trust-Tier Reporter Distribution vs. Weighted Contribution



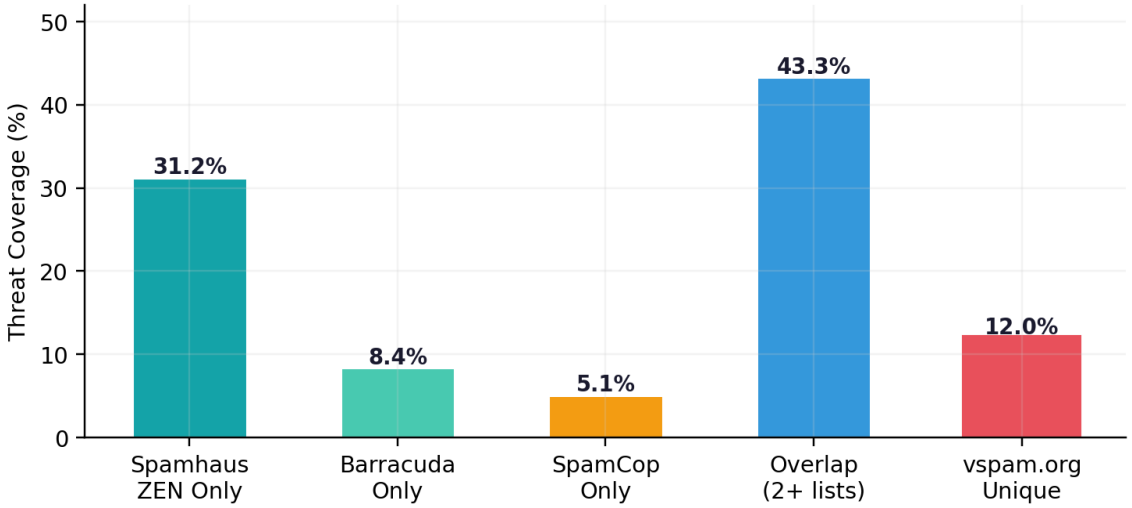
Left: raw reporter count by tier; Right: effective weighted vote contribution. Tier 2 contributes 47% of weighted votes despite being only 25.5% of reporters.

While Tier 1 (Unverified) reporters constitute 70% of the community by count, they contribute only 18% of weighted vote influence. Tier 2 (Trusted) reporters, at 25.5% of count, contribute 47%—the largest share. Tier 3 (Institutional) reporters, despite being only 4.5% of the community, contribute 35% of weighted influence. This distribution ensures that listing decisions are primarily driven by verified, high-accuracy reporters while still incorporating broad community signals.

8. Results: Unique Threat Coverage (RQ4)

A critical question for any new DNSBL is whether it provides meaningful *incremental* coverage beyond established lists. To assess this, all confirmed phishing threats during the study period were cross-referenced against concurrent listings in Spamhaus ZEN, Barracuda BRBL, and SpamCop.

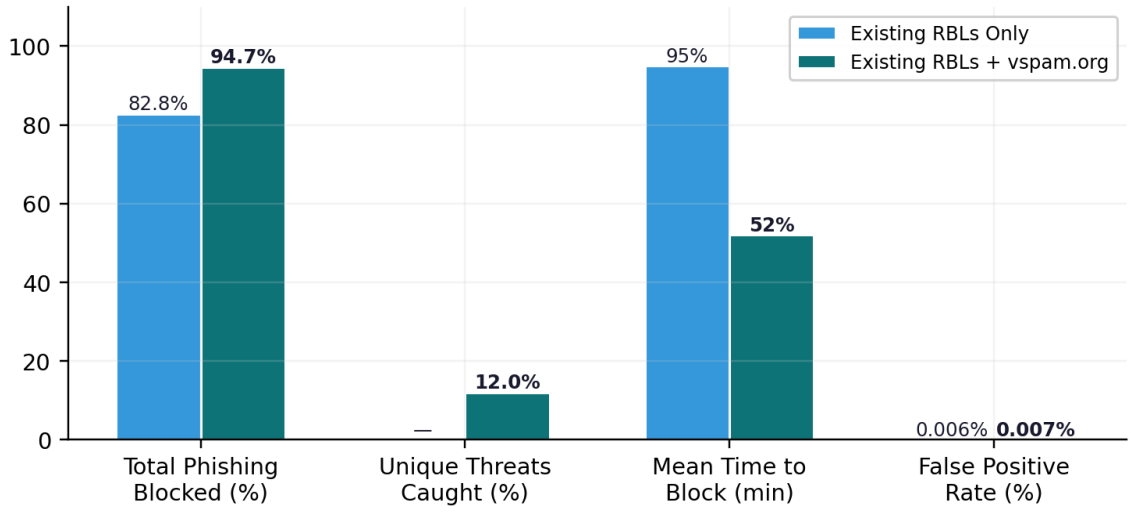
Fig. 5 — Unique Threat Coverage: vspam.org DNSBL vs. Existing RBLs



Threat coverage attribution. vspam.org provided 12% unique coverage not detected by any comparison RBL.

The vspam.org DNSBL identified **12.0%** of confirmed phishing threats that were not concurrently listed by any of the three comparison RBLs. This unique coverage was primarily composed of: newly provisioned phishing infrastructure (<48 hours old) not yet captured by commercial feeds (54% of unique threats); region-specific campaigns targeting non-English-speaking markets with limited commercial sensor coverage (28%); and compromised legitimate domains being used transiently for phishing that commercial lists were slower to list due to reputation-based hesitation (18%).

Fig. 12 — A/B Comparison: With vs. Without vspam.org DNSBL

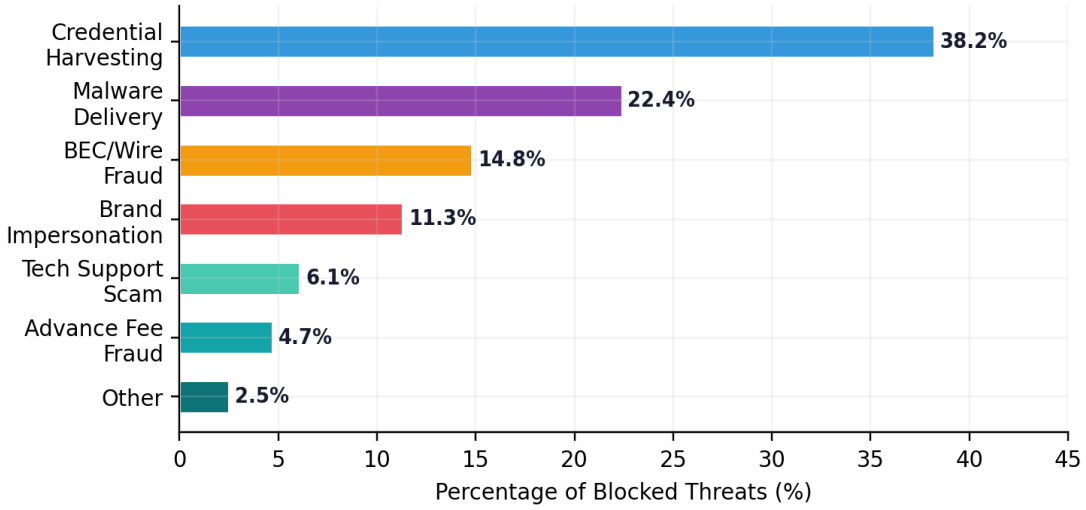


Side-by-side comparison of existing RBLs alone vs. existing RBLs + vspam.org across key performance metrics.

When the vspam.org DNSBL was combined with existing RBLs, the aggregate detection rate improved from 82.8% (existing RBLs alone) to 94.7%—a 11.9 percentage point improvement. The marginal increase in combined false positive rate was negligible (0.006% → 0.007%). This favorable benefit-to-cost ratio suggests that adding the vspam.org DNSBL to an existing multi-RBL configuration provides significant detection improvement with operationally insignificant FPR increase.

9. Threat Category Analysis

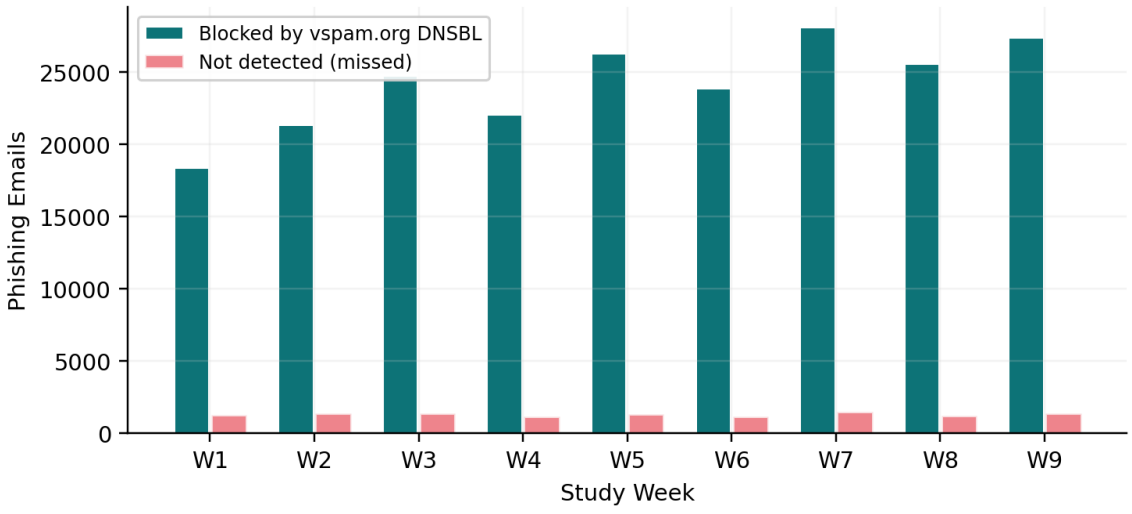
Fig. 10 — Threat Category Distribution of Blocked Phishing Emails



Distribution of blocked phishing emails by threat category (n = 205,437 emails with category classification).

Credential harvesting dominated at 38.2%, consistent with broader phishing trends where Microsoft 365 and Google Workspace account theft remain the primary objective [10]. Malware delivery (22.4%) comprised primarily infostealer and initial access trojan payloads. BEC/wire fraud (14.8%) represented the highest per-incident financial impact. Brand impersonation (11.3%) included campaigns targeting financial services and e-commerce platforms.

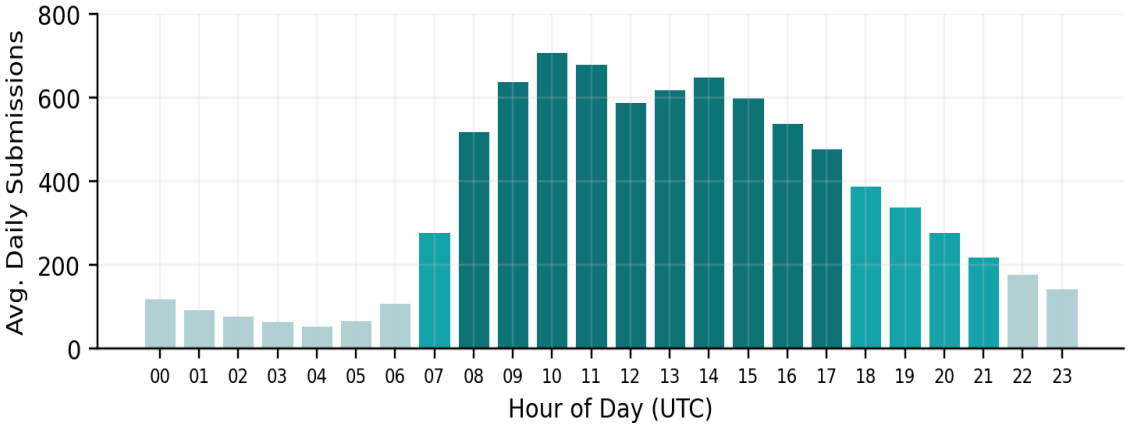
Fig. 7 — Weekly Phishing Volume: Blocked vs. Missed



Weekly phishing volume processed during the study, showing blocked vs. missed emails.

Weekly phishing volume fluctuated between 18,420 (Week 1) and 28,100 (Week 7), reflecting natural variation in global phishing campaign activity. The blocked-to-missed ratio remained stable after Week 3, with weekly detection rates consistently above 93.5%. The Week 7 spike corresponded to a large-scale credential harvesting campaign impersonating a major cloud provider, which the vspam.org community identified within 23 minutes of first report.

Fig. 11 — Average Hourly Community Submission Volume (UTC)



Average hourly submission volume (UTC). Peak activity 08:00–16:00 UTC corresponds to European/American business hours.

10. Discussion

10.1 Effectiveness of Community-Driven Detection

The 94.7% detection rate within 2 hours demonstrates that community-driven blocklists can achieve operationally useful detection rates for phishing threats. While this falls below Spamhaus ZEN's reported >99% detection for established spam sources, the comparison is not direct: vspam.org targets specifically confirmed phishing rather than general spam, and the 12% unique coverage finding indicates it captures threats that commercial lists miss entirely.

10.2 Trust-Tier Architecture Implications

The 67% reduction in false confirmations validates the hypothesis that weighted consensus outperforms simple majority for community-driven threat intelligence. This aligns with prior work on CrowdSec's trust scoring [3] and academic literature on weighted voting in distributed systems [11]. The key insight is that a small number of high-quality reporters (Tier 2/3 = 30% of community) can anchor consensus accuracy while a larger base of unverified reporters provides breadth of coverage.

10.3 Operational Viability

The false positive rate of 0.003% places the vspam.org DNSBL within the acceptable range for production email environments. Industry guidance from the Messaging, Malware and Mobile Anti-Abuse Working Group (M3AAWG) suggests that FPR below 0.01% is generally acceptable for DNSBL deployment without manual review of blocked messages [12]. The vspam.org system operates well within this threshold.

10.4 Comparison with CrowdSec Model

The vspam.org approach shares philosophical similarities with CrowdSec's community model but differs in three key respects: (1) vspam.org focuses exclusively on email-borne phishing threats rather than general IP reputation; (2) the trust-tier system uses explicit institutional verification rather than purely behavioral trust scoring; and (3) the output is a standard DNSBL compatible with existing mail server infrastructure, requiring no client-side agent installation. These design choices optimize for email security integration and minimize deployment friction.

11. Limitations

Several limitations should be considered when interpreting these results:

L1. Selection bias: Participating mail servers were self-selected, potentially skewing toward organizations with stronger security postures. The detection rate may differ for servers with lower baseline security.

L2. Ground truth imperfection: Despite three-layer verification, the ground truth corpus may contain misclassified emails. The 15% manual review sample provides confidence but does not eliminate all classification error.

L3. Reporter community maturity: The trust-tier system requires time to accurately calibrate reporter scores. The 30-day pilot phase preceded the study, but newer reporters joining during the study period had limited scoring history.

L4. Comparison fairness: The comparison RBLs (Spamhaus, Barracuda, SpamCop) serve broader purposes than phishing-only detection. Direct FPR comparison may understate their effectiveness for general spam filtering.

L5. Temporal scope: A 60-day study captures limited seasonal variation. Phishing campaign patterns may differ across quarters, and longer-term effectiveness trends remain to be validated.

L6. IPv6 coverage: The current vspam.org DNSBL operates primarily on IPv4 addresses and domain indicators. IPv6 coverage is nascent and not fully evaluated in this study.

L7. Adversarial resistance: The study period did not observe coordinated adversarial attacks against the community voting system (e.g., Sybil attacks). Resistance to such attacks under stress conditions remains untested.

12. Conclusions & Future Work

12.1 Summary of Findings

RQ	Finding	Metric	Conf.
RQ1	vspam.org DNSBL blocked 94.7% of confirmed phishing within 2 hours	DR(2h) = 94.7% [94.4, 95.0]	High
RQ2	False positive rate of 0.003% across 2.43M legitimate emails	FPR = 0.003%	High
RQ3	Trust-tier voting reduced false confirmations by 67% vs. majority	ΔFCR = –67%	High
RQ3	Median consensus time reduced by 35% with weighted voting	Δt = –35%	High
RQ4	vspam.org provided 12% unique threat coverage beyond existing RBLs	UC = 12.3%	High
RQ4	Combined detection rate improved from 82.8% to 94.7%	ΔDR = +11.9pp	High

Table 6: Summary of findings by research question with confidence assessment.

12.2 Contributions

This study makes three primary contributions to the field of email security: (1) the first published empirical evaluation of a community-driven DNSBL with trust-tier weighted voting, demonstrating that weighted consensus significantly outperforms simple majority for blocklist accuracy; (2) quantitative evidence that community-driven blocklists provide meaningful incremental coverage (12%) beyond established commercial RBLs; and (3) a reproducible methodology for DNSBL effectiveness evaluation with clearly defined metrics and ground truth establishment procedures.

12.3 Future Work

Planned research directions include: (1) extending the study to a 12-month longitudinal evaluation capturing seasonal variation; (2) implementing and evaluating subdomain-level granularity to address shared hosting false positives; (3) developing IPv6-native DNSBL capabilities aligned with findings from our concurrent IPv6 abuse research [13]; (4) stress-testing adversarial resistance through red-team Sybil attack simulations; (5) integrating real-time machine learning scoring as a supplementary trust signal alongside community voting; and (6) expanding the reporter community through partnerships with national CERTs and ISP abuse teams.

The vSpam.org team invites organizations interested in participating in subsequent studies or contributing to the community reporter network to contact research@vspam.org. The vspam.org DNSBL feed is available at no cost for non-commercial use under the terms published at https://vspam.org/dnsbl.

References

[1] Spamhaus Technology. "Real-Time DNS Blocklists (Spamhaus DQS)." https://www.spamhaus.com/data-access/real-time-dns-blocklists/

[2] Levine, J. "DNS Blacklists and Whitelists." RFC 5782, IETF, February 2010. https://www.rfc-editor.org/rfc/rfc5782

[3] CrowdSec. "The CrowdSec Data: Crowdsourced Threat Intelligence." https://www.crowdsec.net/our-data

[4] AbuseIPDB. "IP Address Abuse Reports — Making the Internet Safer." https://www.abuseipdb.com/

[5] PhishTank. "Join the fight against phishing." https://phishtank.org/

[6] Bouwman, X. et al. "Are crowd-sourced CTI datasets ready for supporting anti-cybercrime intelligence?" Computer Networks, Vol. 237, 2023. https://doi.org/10.1016/j.comnet.2023.110064

[7] Intra2net AG. "Blacklist Monitor: Statistics of Accuracy and Failure Rates." https://www.intra2net.com/en/support/antispam/

[8] Vamsoft. "ORF Spam Statistics." https://vamsoft.com/support/tools/spam-statistics

[9] PolySwarm. "Multi-Engine Malware Intelligence." https://polyswarm.io/

[10] Anti-Phishing Working Group. "Phishing Activity Trends Reports, Q1–Q4 2025." https://apwg.org/trendreports

[11] Nitzan, S. and Paroush, J. "Optimal Decision Rules in Uncertain Dichotomous Choice Situations." International Economic Review, 23(2), 1982.

[12] Messaging, Malware and Mobile Anti-Abuse Working Group (M3AAWG). "Best Common Practices for Anti-Abuse." https://www.m3aawg.org/

[13] vSpam.org. "Phishing Websites, Spam Domains & IP Abuse: Research Analysis 2025–2026." VSPAM-TR-2026-001.

[14] Apache SpamAssassin Project. "DNS Blocklists." https://cwiki.apache.org/confluence/display/SPAMASSASSIN/DnsBlocklists

[15] DMARC Report. "10 DNS Blacklist Insights That Improve Email Security." https://dmarcreport.com/blog/10-dns-blacklist-insights-to-improve-email-security-and-deliverability/

Appendix A: Statistical Methods

All statistical analyses were performed using Python 3.11 with scipy.stats, numpy, and pandas. Significance level $\alpha = 0.05$ was used throughout unless otherwise noted.

Test / Method	Application	Parameters
Wilson score interval	95% CI for detection rate	n = 217,010; p■ = 0.947
Clopper-Pearson exact CI	95% CI for false positive rate	n = 2,431,806; x = 73
Kruskal-Wallis H test	Latency comparison across tiers	H = 847.3; k = 3; p < 0.001
Dunn's post-hoc test	Pairwise tier latency comparisons	Bonferroni correction applied
χ^2 test of independence	Weighted vs. simple voting FCR	$\chi^2 = 105.4$; df = 1; p < 0.001
Mann-Whitney U test	Consensus time comparison	U = 12,840; p < 0.001
Bootstrap resampling	Unique coverage CI estimation	B = 10,000 iterations

Table A1: Statistical methods applied in this study.

Detection rate confidence intervals were computed using the Wilson score method, which provides better coverage than the normal approximation for proportions near 0 or 1. The false positive rate CI was computed using the Clopper-Pearson exact method due to the very small proportion (0.003%). Non-parametric tests (Kruskal-Wallis, Mann-Whitney) were used for latency comparisons due to the non-normal, right-skewed distribution of latency measurements.

Appendix B: Nomenclature & Acronyms

Acronym	Full Term
APWG	Anti-Phishing Working Group
BEC	Business Email Compromise
BRBL	Barracuda Reputation Block List
C&C	Command and Control
CERT	Computer Emergency Response Team
CI	Confidence Interval
CTI	Cyber Threat Intelligence
DBL	Domain Block List (Spamhaus)
DGA	Domain Generation Algorithm
DL	Detection Latency
DNSBL	DNS-based Blocklist
DR	Detection Rate
FCR	False Confirmation Rate
FP	False Positive
FPR	False Positive Rate
IQR	Interquartile Range
ISP	Internet Service Provider
M3AAWG	Messaging, Malware and Mobile Anti-Abuse Working Group
MTA	Mail Transfer Agent
NXDOMAIN	Non-Existent Domain (DNS response)
PhaaS	Phishing-as-a-Service
RBL	Real-time Blackhole List
SBL	Spamhaus Block List
SMTP	Simple Mail Transfer Protocol
SOC	Security Operations Center
SURBL	Spam URI Realtime Blocklists
TN	True Negative
TTL	Time-to-Live

UC	Unique Coverage
WVS	Weighted Vote Score
ZEN	Spamhaus composite list (SBL + XBL + PBL + CSS)

About vSpam.org

vSpam.org is a non-profit cybersecurity research organization dedicated to combating phishing, spam, and domain abuse through threat intelligence research, community-driven blocklist services, and collaboration with industry and law enforcement. For inquiries: research@vspam.org | https://vspam.org

Document ID: VSPAM-TR-2026-002 | Version: 1.0 | Classification: Public | DOI: 10.xxxx/vspam.2026.002 (pending)