StrongestLayer

# The Architecture of Evasion

## Why QR Code Phishing Defeats Enterprise Email Security—An Analysis of ~200 Attack Elements Across 19 Campaigns

**StrongestLayer Threat Research**
January 2026

A comprehensive analysis of QR code phishing attacks that bypassed Microsoft Defender, Proofpoint, Mimecast, and Google Workspace controls.

# Executive Summary

Between September 2025 and January 2026, StrongestLayer's threat research team analyzed approximately 200 malicious QR code instances—98 phishing emails containing 106 QR codes, traced through their redirect chains to terminal destinations—that successfully bypassed deployed enterprise email security gateways including Microsoft Defender for Office 365, Google Workspace native controls, Proofpoint, and Mimecast. These attacks represent the most sophisticated and evasive examples of the QR phishing threat: they defeated detection capabilities specifically designed to stop them.

This report examines not whether QR code phishing exists—every security team knows it does—but why it succeeds despite unprecedented industry investment in countermeasures. The answer lies not in detection accuracy, vendor competence, or analyst vigilance. It lies in architecture.

## Key Findings at a Glance

**100%**
QR IN PDF
All attacks used QR codes in PDF attachments

**89%**
FAKE CAPTCHA
Anti-analysis techniques defeated sandboxes

**5x**
GROWTH
Volume grew 5x in three months

**84%**
UNIQUE URLS
Victim email embedded in URL

**THE CORE PROBLEM**

QR code phishing exploits the mobile device gap: attacks succeed on personal smartphones outside the corporate security perimeter, where no email security solution can operate. This is an architectural constraint, not a detection accuracy problem.

## Key Implications for Security Leaders

**For CISOs:** The question is not "which vendor detects QR code phishing better" but "does our security architecture address threats that execute outside our perimeter?" Organizations investing solely in detection improvements are addressing symptoms, not causes. **For Security Architects:** Legacy email security operates on an implicit assumption: threats can be neutralized at the gateway. QR code phishing invalidates this assumption by design. The attack completes on a device the organization doesn't control, in a browser the organization can't inspect, over a network the organization doesn't monitor. **For the Industry:** When state-sponsored actors (Kimsuky/APT43) and commodity cybercriminals achieve identical success rates using identical techniques against identical defenses, the failure mode is architectural—not vendor-specific.

| Finding | Data Point |
|---|---|
| Attacks using QR codes in PDF attachments | 100% (98/98 emails) |
| Attacks using fake CAPTCHA anti-analysis | 89% (17/19 campaigns) |

| | |
|---|---|
| Attacks with unique URLs (victim email embedded) | 84% (16/19 campaigns) |
| Attacks abusing legitimate infrastructure | 42% (8/19 campaigns) |
| Attacks using self-send spoofing | 26% (5/19 campaigns) |
| Average Jaccard similarity (attack diversity) | 0.209 (below 0.30 threshold) |
| High-sophistication campaigns (Tier 1) | 47% (9/19 campaigns) |
| Estimated human engagement rate (Tier 1) | 75-90% |

# Attack Anatomy: A Complete Kill Chain

To understand why QR code phishing defeats enterprise security, we must examine how these attacks actually work. The following case study walks through a representative Tier 1 attack—the enviousenergy.com campaign—from initial delivery to credential theft.

## The enviousenergy.com Campaign

**Sophistication Score:** 14/18 (Tier 1) | **Samples Analyzed:** 6 | **Estimated Engagement:** 85% This campaign targeted employees at multiple organizations with emails purporting to be from internal HR systems regarding December 2025 bonus qualification. The attack was notable for combining three distinct evasion techniques that, in combination, defeated every deployed email security gateway it encountered.

## Attack Sequence

| Stage | What Happens | Why Detection Fails |
|---|---|---|
| 1. Delivery | Email arrives with PDF attachment titled "December_2025_Bonus_Review.pdf" | PDF contains no malicious code—only an image |
| 2. QR Presentation | PDF displays QR code with message: "Scan to view your personalized bonus statement" | URL embedded in image, not extractable by legacy parsers |
| 3. Initial Redirect | QR decodes to google.com/url?q=[encoded destination] | Google.com has pristine reputation; SEGs don't flag |
| 4. Second Redirect | Google redirects to enviousenergy.com (recently registered, clean history) | No negative reputation yet; domain age <30 days |
| 5. Anti-Analysis | enviousenergy.com serves fake CAPTCHA with 8-second timing check | Automated sandboxes timeout or fail CAPTCHA |
| 6. Final Redirect | After CAPTCHA, redirects to payroll.vaicrotea.us | Only reachable after human interaction |
| 7. Credential Harvest | Victim sees Microsoft 365 login page, enters credentials | Occurs on personal mobile device—no corporate visibility |

## Why This Attack Succeeds

Each stage of this attack is designed to exploit a specific limitation in email security architecture: **The PDF Container:** By embedding the QR code as an image within a PDF, the attacker ensures the malicious URL is not present as extractable text. Legacy email security analyzes text content, embedded links, and attachments for known malicious patterns. An image containing a QR code is effectively invisible to these systems. Modern SEGs have added image analysis capabilities, but the QR code merely points to Google—not a malicious destination. **The Reputation Laundering Chain:** The attack uses Google's redirect service as a reputation shield. When the SEG decodes the QR code and analyzes the URL, it sees google.com—a domain with perfect reputation. Even systems sophisticated enough to follow redirects encounter enviousenergy.com, which has no negative history because it was registered specifically for this campaign. By the time threat intelligence flags it, the campaign has moved on. **The Anti-Analysis Gate:** The fake CAPTCHA page serves a critical function: it filters out automated analysis. Sandboxes and crawlers cannot solve the CAPTCHA or satisfy the timing requirement. This ensures that security tools analyzing the attack chain never reach the actual credential harvesting page—and therefore cannot detect or block it.

> **KEY INSIGHT**
> The credential theft occurs on the employee's personal smartphone, in a personal browser, over a personal network. No corporate security control—email gateway, EDR, SIEM, or firewall—has visibility into this transaction. The attack was designed from inception to complete outside the security perimeter.

# The Industry Response and Its Limits

The email security industry's response to QR code phishing has been neither slow nor inadequate by traditional standards. Every major vendor has invested significantly in detection capabilities. Understanding what these investments can and cannot accomplish is essential for realistic security planning.

## Vendor Investment Has Been Substantial

**Microsoft** acknowledged "several months" in 2023-2024 where customers experienced massive increases in malicious QR codes reaching inboxes. By late 2024, Microsoft reported blocking 3 million QR code phishing attempts daily at peak—since decreased to approximately 200,000 attempts per day as attackers adapted techniques. Microsoft Defender for Office 365 now includes QR code extraction from images, URL analysis, and real-time reputation checking. Yet attacks continue to reach protected inboxes.

**Proofpoint** added inline QR sandboxing capabilities in December 2023, marketing "99.99% efficacy" in threat detection. The company identified 4.2 million QR-based threats in H1 2025. Proofpoint's Targeted Attack Protection (TAP) now includes image extraction from attachments, QR code decoding, URL analysis, and redirect chain traversal. Despite these capabilities, Proofpoint-protected organizations appear repeatedly in our dataset.

**Mimecast** deployed QR code analysis at scale, now processing 3.5 million QR codes daily. In a single campaign between April and June 2024, Mimecast detected 70,000 malicious QR codes. Notably, Mimecast recommends setting detection thresholds at 90% probability—explicitly acknowledging a 10% miss rate as operationally necessary to avoid overwhelming analysts with false positives.

**Abnormal Security** released a dedicated QR code parser in late 2023, reporting that 17% of attacks bypassing native email controls contain QR codes, with 89% focused on credential theft. Abnormal's behavioral analysis approach represents a different architectural model than traditional SEGs, yet QR code attacks continue to succeed.

> *The vendor response has been professional and well-resourced. The problem is that detection-based approaches face structural constraints that investment cannot resolve.*
>
> — StrongestLayer Threat Research

**THE CONTRADICTION**

Despite coordinated, well-resourced industry response, successful QR code phishing incidents grew 5x between August and November 2025—from 46,296 to 249,723 in just three months (Kaspersky Labs). Vendor investment is not translating to threat reduction.

# Three Structural Constraints

The attacks analyzed in this report reveal not a detection gap that vendor investment can close, but architectural constraints that affect every legacy email security platform equally. These constraints are structural—inherent to how email security has been designed—and cannot be resolved by improving detection accuracy.

## Constraint 1: The Mobile Device Gap

The most fundamental problem is architectural and affects every vendor equally. Consider the attack sequence that QR code phishing exploits: **Step 1:** Malicious email delivered to corporate inbox (protected environment) **Step 2:** User scans QR code with personal smartphone (unmanaged device) **Step 3:** Malicious website loads in personal mobile browser (no corporate controls) **Step 4:** Credentials entered and stolen (outside security perimeter entirely) Every email security vendor acknowledges this constraint. Corporate email security can analyze the email containing the QR code. It cannot protect what happens after the user points their personal phone camera at their computer screen.

The FBI's January 2026 flash alert directly validates this analysis, characterizing quishing as operating "outside normal Endpoint Detection and Response (EDR) and network inspection boundaries." This is not a failure of specific products—it is a fundamental limitation of perimeter-based security when the attack completes outside the perimeter.

## Constraint 2: Multi-Stage Redirects Defeat Analysis

Multi-stage redirects appeared in **177.9%** of analyzed attacks—meaning the average attack employed 1.78 redirect techniques, with many attacks stacking 2-3 methods simultaneously. This creates an analysis problem that scales exponentially.

**A representative attack chain:** QR code → google.com/url?q=... (Google redirect, pristine reputation) → enviousenergy.com (recently registered, no negative history) → payroll.vaicrotea.us (fake

CAPTCHA with timing validation) → credential harvesting form When SEGs decode the QR code and analyze the initial URL, they encounter Google—with excellent security reputation. Even sophisticated systems that follow redirects often cannot reach terminal destinations when timing checks and fake CAPTCHA pages specifically block automated analysis.

### Legitimate Infrastructure Abuse

| Platform | Usage | Platform | Usage |
|---|---|---|---|
| AWS S3 | 23% | Microsoft Azure | 12% |
| Google (redirects) | 18% | AgileCRM | 5% |
| Cloudflare | 15% | Other legitimate | 27% |

*Blocking these platforms would break legitimate business operations. Attackers deliberately exploit this constraint.*

### Constraint 3: Mathematical Detection Limits

Attack diversity has reached thresholds where pattern-matching faces fundamental constraints. **Jaccard similarity** measures shared characteristics between attacks—a standard metric for evaluating pattern-based detection feasibility. Traditional template-based phishing shows 0.85-0.95 Jaccard similarity (85-95% shared features), making pattern-matching highly effective. QR code campaigns in our dataset averaged 0.209 Jaccard similarity. Targeted campaigns averaged just 0.134—meaning each attack shared less than 14% of its characteristics with other attacks in the same campaign.

Below the 0.30 threshold, pattern-based detection faces a mathematical trap: **Aggressive detection** catches more attacks but generates false positive rates that overwhelm analyst capacity. **Cautious detection** maintains operational feasibility but misses novel attacks entirely. Microsoft's March 2025 acknowledgment that 90-95% of employee-reported "suspicious" emails are false positives reflects this constraint in practice. When the vast majority of flagged messages are legitimate, analysts develop alert fatigue—and real threats slip through.

> **THE BOTTOM LINE**
>
> These three constraints—mobile device gap, multi-stage redirects, and mathematical detection limits—are structural. No amount of detection investment resolves architectural limitations. Organizations must evaluate security solutions based on whether they address these constraints, not whether they promise higher detection rates.

# Campaign Intelligence

Our analysis identified 19 distinct phishing campaigns, each scored across six dimensions of sophistication: three measuring social engineering quality (pretext believability, brand impersonation fidelity, personalization depth) and three measuring technical execution (infrastructure resilience, anti-analysis capabilities, URL obfuscation sophistication).

### SEG Evasion Technique Prevalence

| Evasion Technique | Usage | SEG Limitation Exploited |
|---|---|---|
| QR code in PDF attachment | 100% | URL not extractable from image-based content |
| Victim email embedded in URL | 84% | Every URL unique—defeats IOC matching |
| Fake CAPTCHA anti-analysis | 89% | Sandbox detonation defeated by timing gates |
| Legitimate infrastructure abuse | 42% | AWS/Google/Cloudflare have trusted reputation |
| Self-send spoofing | 26% | Self-addressed email often whitelisted |
| Google redirect abuse | 11% | google.com has pristine reputation globally |
| Subdomain brand impersonation | 32% | Subdomain impersonation often undetected |

## Campaign Distribution by Sophistication Tier

| Tier | Score Range | Campaigns | Samples | % of Attacks | Engagement |
|---|---|---|---|---|---|
| Tier 1 (High) | 12-18 | 9 | 32 | 60% | 75-90% |
| Tier 2 (Medium) | 8-11 | 9 | 19 | 36% | 50-75% |
| Tier 3 (Low) | 0-7 | 1 | 2 | 4% | 25-40% |

**Key Finding:** 60% of attacks with campaign attribution belong to high-sophistication operations, suggesting either well-resourced threat actors or—more likely—LLM-assisted content generation enabling sophisticated attacks at scale. The barrier to creating convincing, professionally-written phishing content has collapsed.

## Selected Tier 1 Campaign Profiles

| Campaign | Score | Samples | Engagement | Primary Technique |
|---|---|---|---|---|
| enviousenergy.com | 14/18 | 6 | 85% | Google redirect + fake CAPTCHA + timing gates |
| jehoosa.my.id | 14/18 | 4 | 90% | Self-send spoofing (sender = recipient) |
| brotli.chioloodro.live | 13/18 | 8 | 80% | Multi-org targeting, 4 distinct sender domains |
| AWS S3 (1978auth007) | 13/18 | 2 | 90% | amazonaws.com trust + self-send + Base64 |
| rippling.triciodou.ru.com | 13/18 | 2 | 85% | HR platform (Rippling) brand impersonation |
| rfraud.netlify.app | 12/18 | 3 | 75% | Netlify hosting + HR document pretext |

# The Human Factor: Why These Attacks Work

Technical evasion explains how QR code phishing defeats email security controls. Understanding human psychology explains why recipients engage with these attacks at rates approaching 90%. The sophistication of social engineering in our dataset was remarkable—and provides crucial context for defensive strategy.

## Self-Send Spoofing: The 90% Engagement Technique

The self-send technique—where the sender email matches the recipient email—achieved the highest human engagement rate in our dataset. This exploits a trust assumption baked into most organizations' security policies. **Why it works:** Organizations frequently whitelist self-addressed email as a productivity feature. Employees routinely send themselves notes, calendar reminders, and document links. Security teams configure policies to allow this traffic without inspection. Attackers exploit this trust by sending emails that appear to come from the recipient themselves. **Example:** An employee at [Company Name] receives an email FROM jsmith@[company].com TO jsmith@[company].com with subject "December 2025 Bonus Qualification - Action Required." Email headers reveal the message originated from infrastructure flagged in threat intelligence databases. But because the email is "from herself," it bypasses security policies designed to scrutinize external senders.

## Human Engagement by Attack Theme

| Attack Theme | Avg Engagement | Attack Theme | Avg Engagement |
|---|---|---|---|
| Self-send + Bonus/Salary | 90% | Salary Amendment Request | 75-80% |
| Performance Review Access | 85% | Voicemail Notification | 70% |
| December/EOY Bonus | 80-85% | Generic Payroll Document | 65-70% |
| HR Platform (Rippling, etc.) | 80-85% | eSignature Request | 65% |

## The Psychology of High-Engagement Themes

The most effective themes share common psychological characteristics: **Personal Financial Stakes:** Bonus and salary themes directly affect the recipient's compensation. The potential loss of expected income creates urgency that overrides caution. Employees don't want to "miss" their bonus because they ignored an HR email. **Temporal Anchoring:** Campaigns referenced actual business cycles—December bonuses, year-end reviews, EOY deadlines. This contextual accuracy makes the emails feel legitimate. An email about "December 2025 Bonus Qualification" arriving in December 2025 matches recipient expectations. **Authority Impersonation:** HR departments, payroll systems, and performance management platforms command compliance. Employees are conditioned to respond promptly to HR requests. Questioning these emails feels like questioning legitimate authority. **Low Cognitive Load:** "Scan this QR code to view your document" requires minimal mental processing. The action is simple, familiar (everyone uses QR codes), and apparently low-risk. Recipients don't pause to evaluate because the requested action seems harmless.

**DEFENSIVE IMPLICATION**

Security awareness training that focuses on identifying suspicious emails misses the point. These emails aren't suspicious—they're professionally crafted to appear completely legitimate. Defense requires architectural solutions that protect users even when they engage with convincing content.

# Nation-State Validation

The FBI's January 2026 flash alert documented active QR code phishing by **Kimsuky** (also tracked as APT43, Velvet Chollima, and Emerald Sleet)—a North Korean state-sponsored cyber espionage group with a decade-long track record of sophisticated operations against government, research, and policy organizations.

**FBI FLASH ALERT — JANUARY 2026**

"Quishing operations frequently end with session token theft and replay, enabling attackers to bypass multi-factor authentication and hijack cloud identities without triggering typical 'MFA failed' alerts."

**Documented Kimsuky QR Phishing Campaigns (May-June 2025):** • Foreign policy questionnaire targeting think tank leaders and academic researchers • Secure document access requests sent to embassy-spoofed recipients • Internal staff communication impersonation targeting government contractors • Conference registration fraud with fake Google login pages **Strategic Significance:** Kimsuky typically employs zero-day exploits, supply chain compromises, and sophisticated malware. The fact that state-sponsored actors with access to elite capabilities choose quishing as a primary vector signals that traditional email security architectures have become reliably exploitable through techniques requiring minimal technical sophistication.

> *For CISOs: if both elite APT groups and commodity cybercriminals succeed with identical techniques against identical defenses, the problem is architectural—not adversary-specific.*
>
> — StrongestLayer Threat Research

## LLM Assistance Indicators

Multiple campaigns in our dataset exhibited characteristics suggesting AI-assisted content generation. The barrier to creating sophisticated, convincing phishing content has effectively collapsed.

**Indicators observed in high-sophistication campaigns:** • **Zero typographical errors** across all Tier 1 samples—unusual for traditional phishing • **Professional HR vocabulary** consistent with genuine corporate communications standards • **Semantic variation** in subject lines: same theme expressed differently across samples within campaigns, suggesting programmatic generation rather than template reuse • **Realistic reference number patterns** (e.g., "REF-2025-DEC-4892") suggesting procedural generation • **Contextually appropriate urgency**: temporal anchoring to actual business cycles (year-end, December deadlines, Q4 reviews) These indicators align with industry findings that 82.6% of

phishing emails analyzed between September 2024 and February 2025 contained AI-generated content (KnowBe4 2025 Phishing Threat Trends Report). The implication: defenders should assume all future attacks will exhibit professional-quality content regardless of attacker resources.

### Emerging Evasion Techniques

Beyond the core techniques documented in this report, our research identified several emerging evasion methods that will likely increase in prevalence:

| Technique | Prevalence | How It Works |
|---|---|---|
| Weaponized Security Language | 8.2% | OAuth/MFA terminology lowers suspicion: "Scan with your authenticated device" |
| ASCII QR Codes | 12% (Jan 2026) | Text characters (■■■) render as QR—no image for OCR to analyze |
| Blob URIs | Emerging | Browser-generated identifiers have no domain for reputation analysis |
| SVG QR Codes | Emerging | Vector graphics evade bitmap-focused image analysis |

## Recommendations for Security Leaders

Organizations that recognize QR code phishing as an architectural problem rather than a detection accuracy problem can evaluate solutions accordingly. The following framework provides concrete criteria for assessing security investments.

### Evaluation Framework for Email Security Architecture

| Capability | Evaluation Questions |
|---|---|
| Evidence Collection | Can it decode QR codes from PDFs? Follow multi-stage redirects to terminal destinations (not just first hop)? Analyze through fake CAPTCHA and timing-check barriers? If any answer is "no," the solution cannot analyze the attacks in this report. |
| Business Context Reasoning | Can it identify sender-content mismatches (Pakistani domain discussing U.S. HR policies)? Recognize organizational relationship violations? Detect temporal incongruity (bonus emails outside bonus season)? |
| Novel Technique Resilience | Does it require explicit programming for each new technique (QR codes, ASCII art, blob URIs), or can it detect malicious intent regardless of construction method? Pattern-matching degrades; reasoning adapts. |
| False Positive Management | What is actual time required to investigate each false positive? At what rate does analyst capacity saturate? If 90%+ of alerts are false positives, detection is creating analyst fatigue without improving security. |
| User Empowerment | Can employees get real-time analysis before engaging with suspicious content? Does the solution address the mobile device gap through informed user decision-making rather than perimeter control alone? |

### Immediate Actions

**1. Audit current QR code detection capabilities** against the evasion techniques documented in this report. Specifically test: Can your SEG decode QR codes from PDF attachments? Can it follow Google redirects to true destinations? Can it reach terminal URLs behind fake CAPTCHA pages? If the answer to any question is "no," you have coverage gaps. **2. Assess false positive operational burden.** Determine if current sensitivity settings create coverage gaps. If 90-95% of flagged emails are false positives, your detection is creating analyst fatigue without improving security posture. Consider whether tuning toward fewer false positives is actually leaving real threats undetected. **3. Implement user-facing analysis tools** that enable employees to evaluate suspicious emails before mobile engagement. The mobile device gap cannot be closed by perimeter controls—but it can be addressed through informed user decision-making. Give employees the ability to check before they scan. **4. Evaluate reasoning-based detection architectures** designed for novel threats rather than pattern-matching against historical attacks. The key question: does this solution require a signature update for each new technique, or can it recognize malicious intent from context? The former will always lag; the latter can adapt.

## Methodology

**Data Collection:** ~200 malicious QR code instances—98 phishing emails containing 106 distinct QR codes—analyzed through complete redirect chains. All attacks bypassed at least one deployed SEG (Google Workspace, Microsoft Defender for Office 365, Mimecast, or Proofpoint) prior to StrongestLayer detection. **Campaign Attribution:** Of 92 production email samples (excluding 6 test records), 53 (57.6%) were attributable to 19 distinct campaigns based on shared destination infrastructure. **Sophistication Scoring:** Six dimensions (0-3 pts each, 18 max): Social Engineering (pretext quality, brand fidelity, personalization) and Technical (infrastructure resilience, anti-analysis, URL obfuscation). Tier 1: 12-18 pts, Tier 2: 8-11 pts, Tier 3: 0-7 pts. **Human Engagement Assessment:** Based on pretext believability, brand alignment, urgency appropriateness, visible red flags, and call-to-action clarity. Estimates validated against industry phishing simulation benchmarks. **Data Limitations:** This dataset comprises attacks that bypassed other controls before StrongestLayer detection—representing the most sophisticated subset of the threat landscape.

## CONCLUSION

The ~200 attack elements analyzed in this report share one characteristic with perfect consistency:
**they bypassed deployed enterprise email security designed specifically to stop them.**

This isn't a detection tuning problem. The mobile device gap creates an architectural constraint
that detection capabilities cannot address. The question isn't whether your current solution can detect QR code phishing. It's whether your security architecture addresses threats that complete outside your perimeter.

# See What Your Current Solution Is Missing

The ~200 attack elements in this report bypassed Microsoft Defender, Proofpoint,
Mimecast, and Google Workspace before detection by StrongestLayer.

Our reasoning-based platform identifies threats that pattern-matching misses—
and empowers employees to make informed decisions before they engage.

**Deploy in 15 minutes. See results in your first week.**

**Request a Consultation**
strongestlayer.ai/demo

StrongestLayer