StrongestLayer

What Your Email Security Can't See

Analysis of 2,042 Advanced Threats That Bypassed Microsoft E3/E5 and Market-Leading Secure Email Gateways

Q3-Q4 2025 | StrongestLayer Threat Intelligence

Executive Summary

This report analyzes what your email security missed.

Every threat in this analysis—all 2,042 of them—successfully bypassed Microsoft Defender (E3/E5) and market-leading secure email gateways before StrongestLayer detected them. This isn't a comparison of detection rates. It's a window into the evasion techniques that define modern email attacks.

The pattern is clear: attackers have stopped trying to look legitimate. Instead, they're hiding behind brands that *already are* legitimate—DocuSign, Microsoft, Google Calendar—platforms so operationally critical that blocking them would halt business.

Key findings

- 77% of attacks impersonated business-critical brands (DocuSign, Microsoft, Google)
- 77% had failed authentication yet reached inboxes—exposing the DMARC enforcement gap
- 17 attacks passed all authentication proving that SPF/DKIM/DMARC validates infrastructure, not intent
- 100% bypassed incumbent security including Microsoft E3/E5 and leading SEGs—by design, we only see what they miss
- ~45% showed AI-assistance markers—a number projected to reach 75-95% within 18 months

Why This Report, Why Now

Email security is experiencing a fundamental inflection point. The techniques that protected organizations for the past decade—pattern matching, reputation scoring, machine learning on historical attacks—are hitting a mathematical wall.

The Pattern-Matching Cliff

Traditional detection relies on similarity. When a new attack resembles previous attacks, signature-based and ML systems can catch it. This worked when attackers reused templates—traditional phishing campaigns show 85-95% similarity across variants, giving detection systems reliable patterns to match.

AI-generated attacks shatter this model. Our Jaccard similarity analysis shows AI-crafted phishing shares only 12-18% common features across variants—each email is effectively unique. When similarity drops below ~30%, pattern-matching becomes mathematically ineffective. We call this the Pattern-Matching Cliff.

The AI Trajectory

Current indicators suggest approximately 45% of sophisticated phishing emails now show markers of AI assistance—grammatically flawless copy, context-aware personalization, and persuasion techniques that exceed typical human attacker capability.

Based on AI adoption curves and attacker economics, we project:

- 75% AI-assisted by end of 2026—AI becomes the default attack authoring tool
- 95% AI-assisted by 2027—manual phishing becomes economically irrational

Organizations relying on pattern-based detection have a narrowing window to augment their defenses with technology designed for novel attacks.

The Trust Exploitation Framework

Attackers have identified a structural vulnerability in enterprise security: platforms too operationally critical to block.

This isn't sophisticated malware or zero-day exploits. It's simpler and more devastating: impersonate the brands that your security team cannot quarantine without breaking the business.

Attack Distribution by Exploited Brand

Exploited Brand	Attacks	% Share	Blockable?
Docusign	*212*	*22.4%*	*No*
Microsoft (M365, Teams, Sharepoint)	187	*19.7%*	*No*
Google Calendar	*68*	*7.2%*	*No**
Financial Institutions	*156*	*16.4%*	*Partially*
Shipping/Logistics (DHL, FedEx, UPS)	*106*	*11.2%*	*Partially*

**Google Calendar attacks bypass email security entirely—invitations arrive via calendar APIs, not email gateways.*

The Authentication Paradox

Email authentication (SPF, DKIM, DMARC) is often positioned as the solution to impersonation attacks. Our data reveals a more complex reality.

The Herd Immunity Problem

77% of attacks in our dataset had failed SPF, DKIM, or DMARC checks—yet still reached recipient inboxes.

Why don't organizations simply reject emails that fail authentication?

The same reason herd immunity requires universal vaccination: DMARC enforcement only works if *everyone* implements it correctly. In practice:

- Many legitimate senders—vendors, partners, clients—still have misconfigured or missing authentication records
- Strict enforcement (p=reject) would block real business emails, damaging operations and relationships
- So most organizations set permissive policies (p=none or p=quarantine) that log failures but don't block
- Attackers exploit this gap—knowing authentication will fail but delivery will succeed

Unlike biological herd immunity, there's no threshold percentage that makes enforcement safe. Even one critical vendor with broken authentication forces you to stay permissive.

When Authentication Passes—And Attacks Succeed

On the other side of the paradox: **17 attacks in our dataset passed all three authentication checks yet were demonstrably malicious.**

These attackers used compromised legitimate infrastructure or exploited platform features like Microsoft 365 Direct Send to send authenticated malicious email. The authentication protocols worked perfectly—validating that the email came from the infrastructure it claimed to come from.

The problem: authentication validates **infrastructure**, not **intent**. It answers "did this email come from where it claims?" not "is this email trying to harm you?"

Email authentication (SPF, DKIM, DMARC) is often positioned as the solution to impersonation attacks. Our data reveals a more complex reality.

Threat Spotlight: Three Campaigns

The DocuSign Campaign (212 Attacks)

The most prevalent attack pattern exploits a simple truth: legal, financial, and healthcare professionals cannot ignore document signing requests. Deal timelines, regulatory deadlines, and court filings create urgency that attackers weaponize.

Attack mechanics: Pixel-perfect DocuSign impersonation → Fake document preview page → Microsoft 365 credential harvesting. The victim believes they're signing a contract; they're surrendering their email password.

Why SEGs fail: URLs point to newly-registered or compromised sites with clean reputation. Multi-stage redirects hide final destination. No malicious payload to scan—pure social engineering.

Sector concentration: 72.5% targeted legal services, where DocuSign is embedded in daily workflow.

The Calendar Exploit (68 Attacks)

This emerging vector bypasses email security entirely. Calendar invitations arrive through calendar APIs, not email gateways—your SEG never sees them.

Attack mechanics: Fake meeting invitation with malicious link in meeting details or location field. By default, invitations auto-populate on victim calendars, creating persistent attack surface without any action required.

Why this matters: As email defenses improve, attackers migrate to less protected channels. Calendar is the canary in the coal mine for Teams, Slack, and other collaboration platform attacks.

The Authentication Bypass (17 Attacks)

These attacks passed every authentication check—SPF, DKIM, and DMARC—by exploiting legitimate infrastructure.

Techniques observed: Compromised legitimate sending infrastructure, M365 Direct Send abuse (sending unauthenticated email via tenant's Direct Send endpoint), and purchased access to authenticated email environments.

Implication: Authentication is necessary but insufficient. Organizations implementing strict DMARC gain protection against spoofing—but not against attackers who've invested in legitimate-appearing infrastructure.

Why Legacy Security Architectures Fail

The attacks in this report share a common characteristic: they don't trigger the signals that traditional email security was built to detect.

The Prosecutor-Only Problem

Both first-generation (pattern matching) and second-generation (ML/statistical) email security systems operate as prosecutor-only architectures. They hunt for evidence of guilt: malicious URLs, suspicious attachments, known-bad patterns, anomalous sender behavior.

But they have no mechanism to prove innocence—to validate that a DocuSign notification is legitimate, that a calendar invitation aligns with real business activity, that a Microsoft alert reflects actual account behavior.

This creates an unsolvable tension:

- Aggressive prosecution: More false positives, blocking legitimate business communications, overwhelming SOC teams
- Conservative prosecution: More false negatives, allowing sophisticated impersonation attacks through

When 77% of attacks exploit trusted brands, this trade-off becomes impossible. Organizations cannot accept the false positive burden of blocking DocuSign, nor the breach risk of letting impersonation attacks through.

The Dual-Evidence Alternative

Detecting trust exploitation attacks requires architectural change: systems that collect both prosecutor evidence (threat indicators) and defender evidence (business legitimacy signals), with reasoning that weighs both.

The key insight: business communication patterns don't change when attack methods evolve. The CFO still uses the same approval workflows. Vendors still follow established procurement processes. DocuSign notifications from legitimate signers still correlate with actual business activity. These legitimacy patterns provide the defender evidence that prosecutor-only systems cannot collect.

Recommendations for Security Leaders

The attacks documented in this report represent a structural challenge, not a configuration gap. Incremental tuning of existing solutions cannot address architectural limitations. However, organizations can take immediate steps to reduce exposure:

Audit your "unblockable" brands. Identify the platforms your organization cannot function without—DocuSign, Microsoft, your industry-specific tools. These are your highest-risk attack surfaces and require detection capabilities beyond pattern matching.

Implement DMARC enforcement progressively. Move from p=none to p=quarantine to p=reject on your own domains. Accept that enforcement gaps in your ecosystem mean authentication alone won't solve impersonation—but it raises the attacker's cost.

Extend visibility beyond email. Calendar invitations, Teams messages, and collaboration tools bypass traditional email security. The calendar exploit (68 attacks) previews where attackers are heading as email defenses improve.

Evaluate reasoning-based detection. Ask vendors: "How do you distinguish a legitimate DocuSign notification from a perfect fake?" If the answer relies on patterns, signatures, or reputation, you have a gap that will widen as AI-generated attacks scale.

Recalibrate user training. Stop training employees to "spot suspicious emails"—these attacks don't look suspicious. Train verification behaviors: out-of-band confirmation for sensitive requests, direct navigation instead of clicking links, healthy skepticism of urgency.

Benchmark false positive rates. If your email security generates 5-15% false positives (industry average), you're burning analyst hours on non-threats while real attacks slip through. A 1% false positive rate saves ~150 analyst hours per month per 1,000 mailboxes.

Methodology

Data Collection

This report analyzes 2,042 confirmed threats detected between September and November 2025 across enterprise environments ranging from 1,000 to 20,000 mailboxes. All data has been anonymized for customer confidentiality.

Detection Architecture

StrongestLayer operates as a supplementary detection layer, analyzing emails that have already passed through incumbent security (Microsoft Defender E3/E5 and/or third-party secure email gateways). By design, 100% of threats in this report bypassed existing enterprise email security.

Detection uses dual-evidence architecture with LLM-based reasoning—collecting both threat indicators and business legitimacy signals to make confident decisions without the false positive burden of prosecutor-only approaches.

Jaccard Similarity Analysis

Attack uniqueness was measured using Jaccard similarity index, comparing feature overlap across attack variants (subject line tokens, body content, sender patterns, URL structures, impersonated brands). Traditional template-based phishing shows 85-95% similarity across campaigns; AI-generated attacks show 12-18% similarity, indicating near-unique generation per target.

Limitations

This analysis covers email-borne threats only. It does not include network intrusions, endpoint attacks, or threats delivered through non-email channels (except calendar invitations). The dataset represents organizations with existing enterprise email security, which may differ from environments without baseline protection.

About StrongestLayer

StrongestLayer provides AI-native email security using dual-evidence reasoning architecture. Purpose-built for organizations seeking to close the detection gap between what legacy systems see and what actually threatens the business.

Contact: research@strongestlayer.com | www.strongestlayer.com