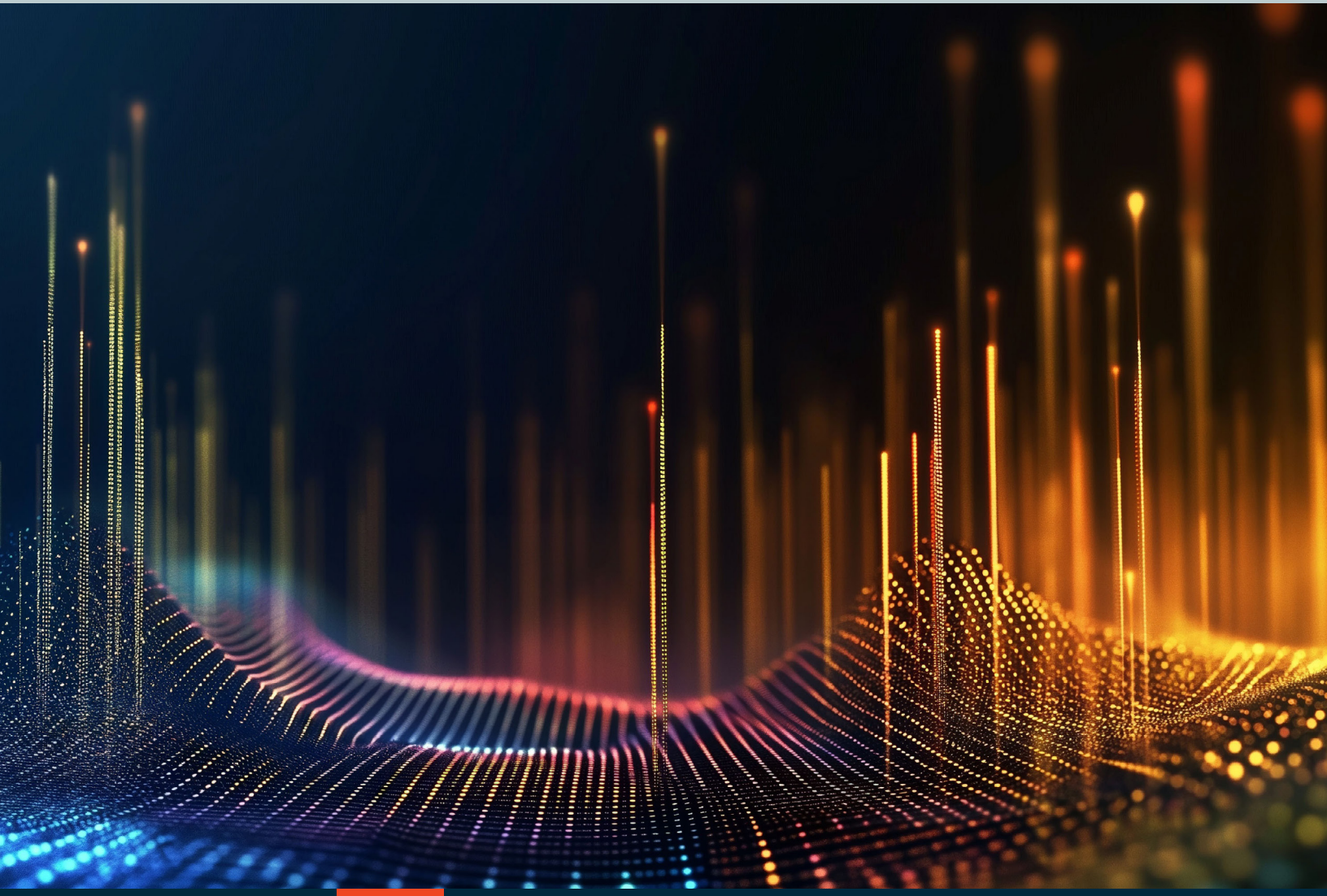
knowbe4

Why Email Security Is Failing—and How to Close the Gap in the Age of AI-Powered Phishing



WWW.KNOWBE4.COM

KnowBe4, Inc. | 33 N Garden Ave, Suite 1200, Clearwater, FL 33755
Tel: 855-KNOWBE4 (566-9234) | Email: Sales@KnowBe4.com

CR0496 0626

Introduction

Artificial intelligence (AI) has fundamentally changed the phishing landscape. Attackers can now craft personalized, context-aware emails that exploit human trust rather than technical vulnerabilities. As a result, many of the tools organizations still rely on to protect their inboxes are no longer sufficient.

Traditional security email gateways (SEGs) were designed to stop bulk spam and known malware. Today, they routinely miss payload-less attacks such as business email compromise (BEC), vendor email compromise (VEC) and advanced phishing techniques that bypass signature-based detection and authentication controls. Even legacy cloud email security platforms—often viewed as a modern replacement for SEGs—are increasingly being outpaced by AI-powered social engineering attacks that blend seamlessly into normal business workflows.

Real-world data confirms this shift. Analysis from KnowBe4 PhishER Plus shows that millions of malicious emails successfully bypass leading email security products every month. The result is higher risk, slower response, increased operational burden and financial damage.

This paper examines why legacy email security models are failing, using empirical data from PhishER Plus to quantify the gap between attackers and defenses. It then outlines a more resilient approach: an integrated strategy that combines cloud email security (CES) with human-vetted intelligence and orchestrated incident response.

Why Security Email Gateways Have Fallen Behind

For years, SEGs have been a foundational layer of email defense. They were built for a threat dominated by bulk spam campaigns, known malware signatures and malicious attachments. In that world, scanning inbound messages for suspicious payloads worked reasonably well. But modern email threats no longer play by those rules. Today's attackers are stealthy, adaptive and AI-powered. As a result, traditional SEGs are falling dangerously behind.

The Shift to Payload-less, Trust-Based Attacks

Modern email attacks are designed to exploit human trust rather than technical vulnerabilities. Instead of relying on obvious malware, attackers favor techniques that leave little to no malicious footprint for a gateway to detect:

- **Business email compromise (BEC)** impersonates executives or trusted partners to pressure employees into wiring funds or sharing sensitive data
- **Vendor email compromise (VEC)** leverages hijacked supplier accounts to submit fraudulent invoices or change payment instructions
- **Account takeover (ATO)** attacks originate from compromised internal mailboxes, making malicious emails appear completely legitimate
- **Advanced phishing techniques** such as HTML smuggling, QR-code phishing ("quishing"), and the hijacking of trusted hyperlinks evade signature-based detection.
- **Polymorphic malware** continuously changes its structure, slipping past static rules and reputation-based controls
- **Identity Management:** Attackers steal or abuse valid credentials, session tokens or permissions to gain legitimate access, allowing them to operate as trusted users, bypass authentication controls and move laterally without delivering malware

These attacks routinely pass authentication checks, such as DMARC, because they often originate from trusted or compromised sources. Since SEGs are optimized to inspect messages before they enter the cloud email environment, they lack the behavioral and contextual insight needed to identify subtle anomalies— such as an unusual payment request, a sudden change in tone or a deviation from established workflows.

The result is predictable. Organizations see more user-reported phishing emails that were never blocked, while security teams spend increasing amounts of time remediating BEC and VEC incidents after the damage is already done.

Data Confirms the Gap

Real-world data underscores just how often SEGs miss modern threats. Over an eight-month period, KnowBe4 PhishER Plus analyzed 37.6 million user-reported emails that had already passed through 22 different SEG and cloud email security platforms. On average, **37% of those reported emails were confirmed threats**—not spam, not false positives, but real attacks.

Using a combination of user reporting and human-validated machine learning, PhishER Plus automatically tagged **14.1 million phishing attempts** during that timeframe. Every single one of those emails made it past a leading email security platform. That translates to nearly **58,000 missed threats per day.**

Additional KnowBe4 Defend data shows that traditional SEGs are now missing **up to 300% more phishing attacks per 10,000 emails** compared to earlier in the year—a clear indicator that attackers are outpacing gateway-based defenses.

Architectural Limitations in a Cloud-First World

The core weakness of the SEG model is architectural. By operating outside the cloud email tenant, SEGs lack access to critical intelligence such as internal mail flow, communication graphs and user behavior baselines. More than half of advanced attacks now originate from compromised accounts—exactly the scenario where reputation-based and perimeter-focused defenses struggle most.

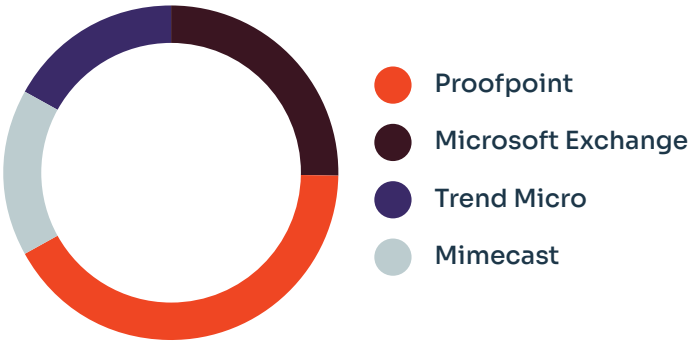
At the same time, running a SEG alongside platforms like Microsoft 365 creates operational drag: duplicated policy engines, multiple quarantine queues and separate admin consoles. This complexity slows response times, increases misconfiguration risk and places unnecessary strain on already overworked security teams.

The takeaway is clear: SEGs were built to stop yesterday's threats. Today's AI-powered email attacks demand a fundamentally different, integrated approach to detection and response.

Not All Cloud Email Security Products Are Created Equal

While cloud-native email security tools offer advantages over perimeter-based SEGs, not all cloud email security products are created equal. Many legacy cloud offerings still rely on detection models rooted in signatures, static rules and post-delivery remediation—approaches that struggle against today's AI-powered, socially-engineered attacks.

Data from KnowBe4 PhishER Plus illustrates this gap clearly that even widely deployed cloud tools are routinely missing a significant volume of real threats. Over the same period, **Microsoft Exchange allowed more than 16 million emails to pass through,** with **35%—nearly 5.8 million—confirmed as malicious. Proofpoint allowed 4.8 million emails through,** of which **57% (2.7 million)** were validated threats. **Mimecast passed 2.1 million emails,** with **22% (477,000)** identified as malicious, while **Trend Micro allowed over 250,000 emails, 23% (59,000)** of which were confirmed attacks.



Simply moving email security to the cloud is not enough. Without advanced detection, automation and integrated incident response working together, legacy cloud email security tools face the same fate as traditional SEGs—falling behind an adversary that is evolving faster than static defenses can adapt.

Why a Comprehensive, Integrated Email Defense Strategy Matters

No single control can stop modern email threats on its own. Effective protection requires a comprehensive strategy that combines **cloud-native email security** with **human-vetted intelligence and incident response.** When these capabilities work together—rather than in isolation—organizations gain the visibility, speed and accuracy needed to stop advanced phishing attacks.

Cloud Email Security: Strong Prevention at the Front Door

A modern CES platform provides critical first-line defense. At a minimum, it must deliver advanced phishing detection and remediation, business email compromise protection, and malware and ransomware defense. These controls reduce overall attack volume and stop many known and emerging threats before they ever reach the inbox.

But prevention alone is not enough. Email is also a major channel for data loss, making outbound security essential. Effective CES includes email DLP, encryption and secure messaging, and protection against misdirected emails.

Strong visibility and management capabilities—such as a unified admin console, detailed reporting and forensics and end-user controls—ensure security teams can understand risk and enforce policy at scale.

Cloud-native architecture is equally critical. Deep integration with cloud email environments provides low latency, full visibility, and access to the telemetry required for accurate detection. Integration with threat intelligence feeds and the broader security ecosystem allows CES to continuously adapt as attacker techniques evolve.

Incident Response: Closing the Loop with Human-Vetted Intelligence

Even the most advanced CES platforms will miss some attacks. That's where an anti-phishing incident response and security orchestration platform like PhishER Plus becomes essential.

PhishER Plus enables users to report suspicious emails with a single click. Those reports are automatically triaged using AI-driven analysis that categorizes messages as clean, spam or threats—prioritizing high-risk phishing attempts and removing guesswork for analysts. Automation then takes over at scale, clustering related messages into campaigns, removing malicious emails from every mailbox with a single action, and isolating similar, unreported threats before they're opened.

Routine tasks—blocking indicators of compromise, opening tickets, updating SIEMs and handling benign reports—are automated, dramatically reducing response times and analyst workload. Just as importantly, users receive feedback on what they reported, reinforcing awareness and turning employees into an active detection layer.

Better Together: Integrated Defense and Measurable ROI

The real value emerges when incident response and CES operate as an integrated system. PhishER Plus feeds human-validated intelligence back into KnowBe4's Cloud Email Security, strengthening detection through continuously improving, AI-driven blocklisting. Crowdsourced intelligence from millions of real-world phishing reports enables proactive protection against emerging campaigns.

The result is a closed-loop defense: **detect, report, analyze, remediate and improve.** By uniting CES with AI-backed, human-vetted incident response, organizations reduce successful attacks, accelerate remediation and transform email security from a reactive cost center into a resilient, ROI-driven defense.

Learn More About

KnowBe4 PhishER Plus

Learn More About

KnowBe4 Cloud Email Security

Free Phishing Security Test
Find out what percentage of your employees are Phish-prone with your free Phishing Security Test

Free Email Exposure Check
Find out which of your users emails are exposed before bad actors do

Free Automated Security Awareness Program
Create a customized Security Awareness Program for your organization

Free Domain Spoof Test
Find out if hackers can spoof an email address of your own domain

Free Phish Alert Button
Your employees now have a safe way to report phishing attacks with one click

About KnowBe4

KnowBe4 empowers the modern workforce to make smarter security decisions every day. Trusted by more than 70,000 organizations worldwide, KnowBe4 is the pioneer of digital workforce security, securing both AI agents and humans. The KnowBe4 Platform provides attack simulation and training, collaboration security, and agent security powered by AIDA (Artificial Intelligence Defense Agents) and a proprietary Risk Score. The platform leverages 15-years of behavioral data to combat advanced threats including social engineering, prompt injection, and shadow AI. By securing humans and agents, KnowBe4 leads the industry in workforce trust and defense.

More information at www.KnowBe4.com.

Follow us on Linkedin

Follow us on X

knowbe4

KnowBe4, Inc. | 33 N Garden Ave, Suite 1200, Clearwater, FL 33755
855-KNOWBE4 (566-9234) | www.KnowBe4.com | Sales@KnowBe4.com