knowbe4

April 2026 | Vol. 7

Phishing Threat Trends Report

How AI is Redefining the Phishing Frontier

2026 is off to a relentless start. The threat landscape is maturing rapidly, bringing new actors, novel tactics, and classic attacks supercharged by AI at scale.

Our 7th Phishing Threat Trends Report delivers actionable intelligence on this evolution. Modern phishing has transcended simple social engineering; today's attackers ruthlessly exploit platform trust and manipulate systems to guarantee a return on investment.

This report dives deep into the diverse approaches of advanced persistent threats (APTs) and criminal groups. We explore the rise of multi-channel attacks targeting collaboration tools, and the mainstream adoption of Adversary-in-the-Middle (AiTM) and reverse proxies in criminal toolkits.

The message is clear: phishing in 2026 is disciplined, persistent, and increasingly AI-enabled. We hope these insights help you navigate this shifting frontier.

Unless otherwise stated, all statistics have been generated from KnowBe4's Collaboration Security products. As always, please reach out if you have any questions or want to learn more about how Knowbe4 stops these threats.

Jack Chapman

SVP of Threat Intelligence, KnowBe4

What's Inside

Opening Stats

Attack Themes Overtime

October 2025

PayPal was mentioned

237.9%

more in phishing emails compared to October 2024

November

Black-Friday e-commerce-based

attacks weaponizing discount codes and time pressure

December

Missed message attacks, specifically utilizing phone numbers

6.2%

of phish in December had a phone number as their payload

January 2026

HR impersonation phish, with

31.0%

of all phish mentioning salary increase/promotions/employee handbook review

February

Zoom and Docusign overtook Microsoft

as the most impersonated brand. How long will that last?

March

36.8%

of malicious HTML attachments

pretended to be an encrypted file in March

17.1% Increase

in phishing attachments compared to previous six months

Average Phishing Campaign Size

2025

22.1 Emails

2026

22.6 Emails

The Average Phishing Link Age

6,372

Days

The Average Domain Age

6,568

Days

The Rising Trajectory of BEC Attacks

Increasing year-over-year

27.0%	2022
39.0%	2023
43.8%	2024
59.1%	2025
61.2%	2026

Payload Distribution

Links (including QR codes)

60.1%

Attachments

30.6%

Social Engineering

9.3%

Opening Stats (continued)

Phishing From Open Source Intelligence (OSINT) Data Breach Distribution

84.3%

Most Common Words in The Subject of Phish

Review
Urgent
Remuneration
Payment
Important

Most Phished Departments

C-Suite
HR
IT
Finance
Sales

Most Phished Industries

Finance
Legal
Healthcare
Logistics
Insurance

Polymorphic Phish

20.9% 2023
26.8% 2024
32.4% 2025
35.6% 2026

Top 5 Legitimate Platforms Facilitated to Deliver Phish

PayPal
Google
Docs, Drive, Classroom
Microsoft
zoom
docusign

Top 10 Attack Techniques in Q1 2026

1. Internal team impersonation
   - 30.0% total attacks
   - 69.5% increase from 2025 ⊕
2. Isolated external link
   - 22.0% total attacks
   - 15.3% increase from 2025 ⊕
3. Image is a clickable link
   - 21.7% total attacks
   - 7.3% increase from 2025 ⊕
4. Date or time in subject line
   - 21.7% total attacks
   - -5.5% decrease from 2025 ⊖
5. Suspicious or mismatched return path
   - 19.6% total attacks
   - 77.0% increase from 2025 ⊕
6. Suspicious attachment file type
   - 19.4% total attacks
   - -3.3% decrease from 2025 ⊖
7. Polymorphic subject line
   - 18.8% total attacks
   - -8.0% decrease from 2025 ⊖
8. Link text is all capitalized
   - 16.5% total attacks
   - 35.5% increase from 2025 ⊕
9. User name in subject line
   - 16.3% total attacks
   - -0.4% decrease from 2025 ⊖
10. Finance-related subject line
    - 14.9% total attacks
    - -5.6% decrease from 2025 ⊖

Attacker Attribution: Unmasking the Adversary

Cybersecurity professionals often struggle to address two critical questions: who is attacking my organization, and why?

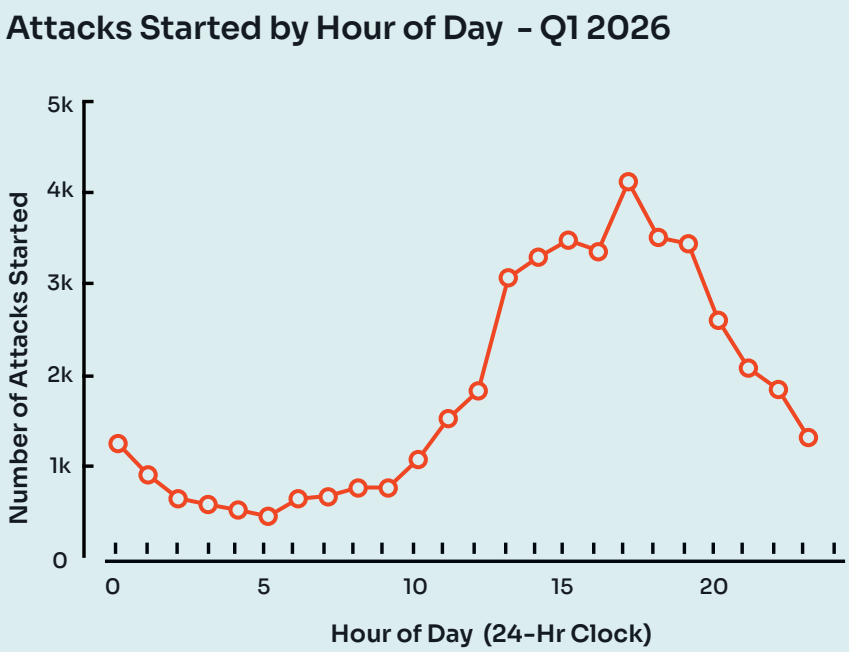
Answering these questions is the foundation of proactive threat intelligence, allowing defenders to anticipate campaigns rather than just react to them. Yet, achieving this in the email space is notoriously difficult. Unlike malware analysis, which leaves concrete forensic artifacts, email is inherently transient. Attackers constantly rotate spoofed domains, sender IPs, and burner accounts to remain anonymous.

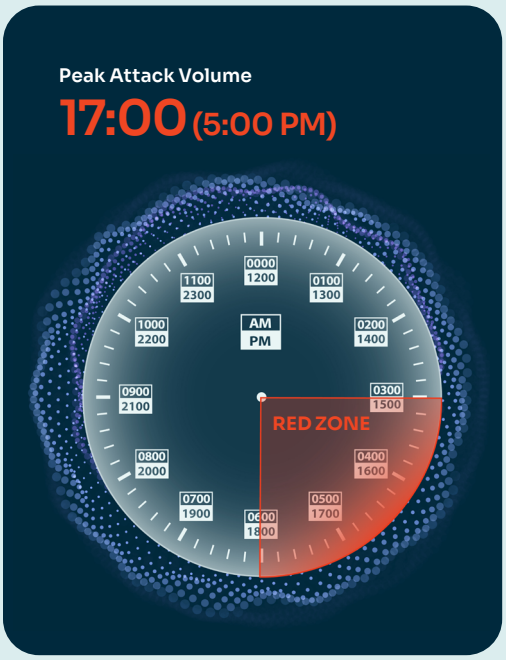
To overcome this, we have moved beyond tracking easily changed indicators and shifted our focus to the attackers' actual behaviors. We combine a broad range of unique intelligence and detection data to provide enhanced visibility into adversary operations, basing our attribution on how these groups operate rather than on perishable data. This approach provides insight into geographic origins, industry targets, and end-to-end evasion strategies that bypass traditional defenses.

Applying this attribution methodology to the 2026 landscape brings the modern threat ecosystem into sharp focus. The following intelligence reveals exactly how adversaries operate at scale — from broad operational trends down to the precise tactics used by some of 2026's most active Advanced Persistent Threats (APTs).

Timing Is Everything

Examining these broad trends begins with the realization that modern threat actors operate on standard corporate schedules. The outdated stereotype of attacks occurring randomly in the middle of the night has been replaced by calculated precision. Today's adversaries strategically align their campaign launches with the natural ebb and flow of the business day.





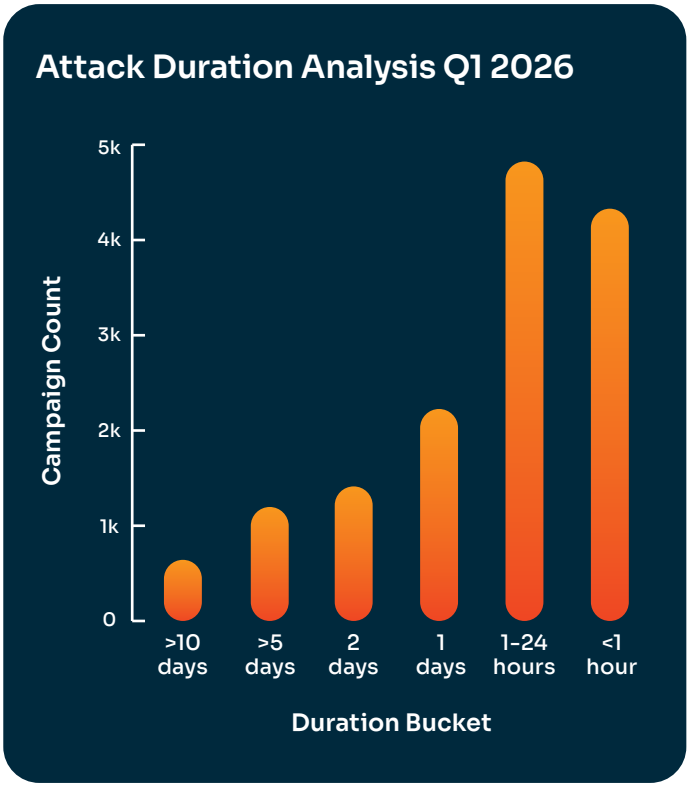
Ask The Expert

James Dyer
Head of Threat Intelligence

We are often asked, "When do people get attacked?"

- We noticed a distinct increase in the number of attacks starting after lunch, with the volume beginning to surge around 1:00 PM (13:00) and accelerating rapidly until the peak at 5:00 PM (17:00) with 4,119 attacks started. The volume then gradually decreases through the late evening and overnight. This trend supports the insight that attackers are deliberately timing campaigns to target employees when they are likely fatigued and less scrutinizing as they wrap up their workday.

While launch timing is designed to exploit human fatigue, campaign duration reveals an adversary's technical evasion strategy. The 2026 data highlights a clear operational divide: rapid execution versus prolonged, sustained operations. While over 60% of campaigns attempt to overwhelm initial defenses in under 24 hours, the most sophisticated APTs are deliberately adopting a sustained approach. By stretching their attacks across five to ten days, these groups use a drip-feed methodology specifically engineered to bypass volumetric security filters and exhaust defense teams.



Attackers are timing their strikes for the "End-of-Day Blur," catching employees when cognitive load is highest and scrutiny is lowest.

Threat Actor Analysis

The true value of behavioral attribution lies in its ability to separate these sophisticated adversaries from the background noise. While our researchers actively monitor over 3,000 unique threat actors, applying this methodology to the 2026 landscape isolates the specific entities driving the macro trends outlined above. **The Threat Actor Table** below profiles a curated selection of the most active APTs identified this year.

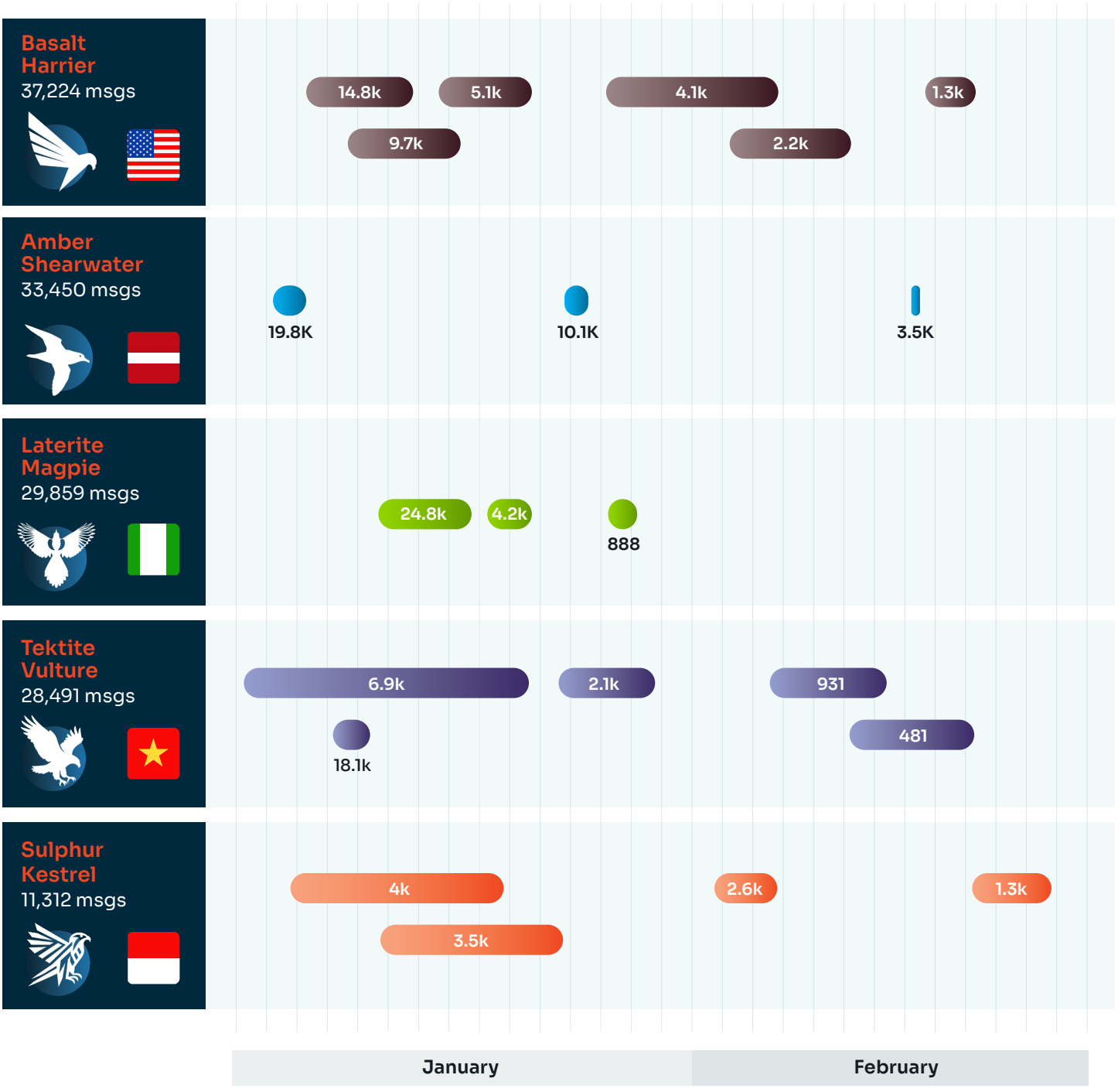
Highlighting this operational shift are groups like Basalt Harrier and Amber Shearwater. Exhibiting near-perfect behavioral consistency, these actors demonstrate immense operational maturity. They have moved beyond manual experimentation, instead aggressively scaling proven, automated models to execute complex use cases like e-commerce fraud at an industrial scale.

Threat Actor Table Q1 2026

Alias	Email Volume	Campaigns	TTP Consistency	Country of Origin	Attack Type	Top Targeted Industries
Basalt Harrier	37224	294	100%	USA	Generic E-commerce Fraud	Finance, Insurance, IT
Amber Shearwater	33450	250	96.8%	Latvia	Temu E-commerce Fraud	IT, Healthcare, Finance
Laterite Magpie	29859	78	100%	Nigeria	BEC + Microsoft Impersonation	Real Estate, Healthcare, Insurance
Tektite Vulture	28491	461	100%	Vietnam	BEC + Docusign Impersonation	Finance, Real Estate, Retail
Sulphur Kestrel	11312	254	100%	Indonesia	BEC + QR Code + Company Impersonation	Insurance, IT, Finance

Understanding an adversary requires seeing their campaigns in motion. The operational timeline below maps the top five APTs, highlighting a clear operational divide: Amber Shearwater executes high-velocity, short-duration strikes, whereas Sulphur Kestrel leverages prolonged, sustained operations to silently bypass security filters. Furthermore, tracking groups like Basalt Harrier reveals densely overlapping campaigns, indicating access to massive, enterprise-grade infrastructure.

Concurrent Campaign Analysis & Geographical Origin



Unmasking APTs: Tektite Vulture and Sulphur Kestrel

To illustrate the distinct tactical differences between high-volume bursts and sustained evasion, we have detailed the operational playbooks of two highly active APTs on the following two pages.

Threat Group Name: Tektite Vulture

Primary Motivation: Financially motivated, specifically Credential Harvesting and Invoice Fraud (Business Email Compromise/BEC).

Origin: Vietnam

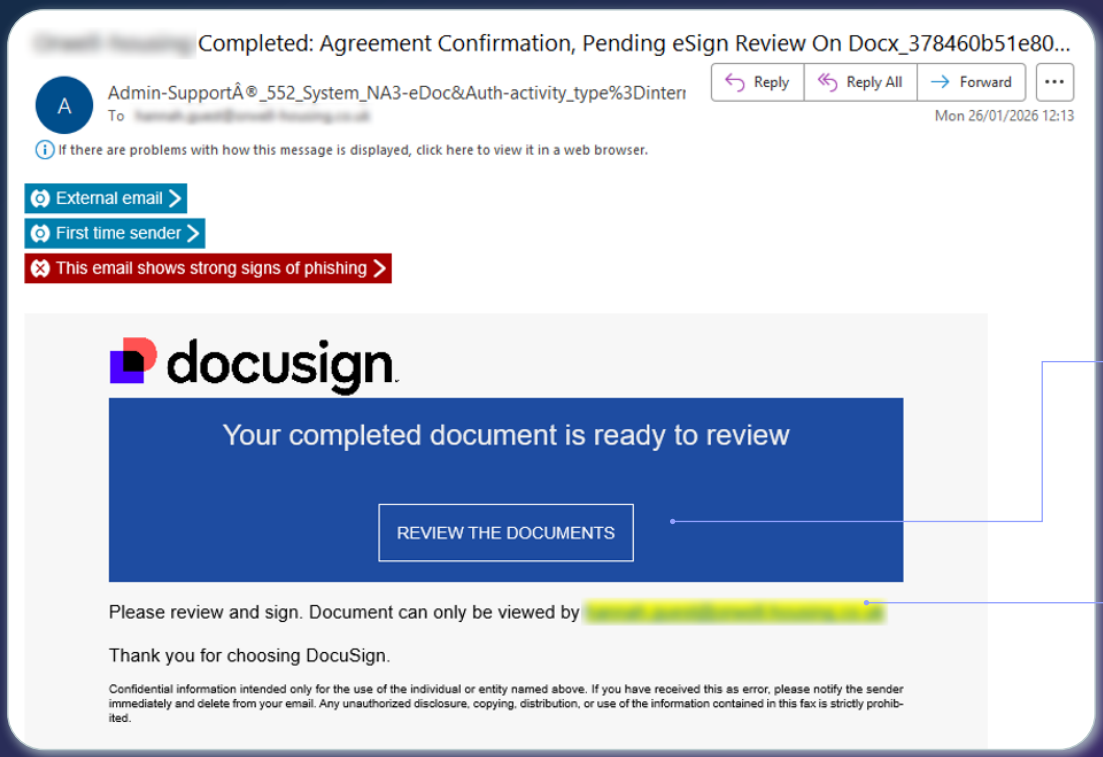
Targets: Large, indiscriminate, and global campaign targeting the private sector across all major geographic regions.

Volume and Tempo: Bad actor operating in large, concentrated bursts. One campaign accounted for 28% of their total volume (8,037 emails) in a single day, designed to overwhelm incident response teams.

Key Tooling: Validated use of the Typhoon Phishing Kit, a Phishing-as-a-Service (PhaaS) toolkit, which allows for highly automated operations and 100% consistency across hundreds of campaigns without manual intervention.

Signature Tactics, Techniques, and Procedures (TTPs):

- **Social Engineering Hook:** Posing as internal IT support or sending fake DocuSign requests to exploit "implicit trust" users have in internal workflows.
- **Evasion:** Aggressively using link obfuscation techniques, including a tactic where the entire email is a clickable image to neutralize text-based NLP filters.
- **Deception:** High usage of "Display Name Stuffing" to push the actual sender's address off-screen on mobile devices (tracked via unusually long sender display names and email prefixes), as displayed in the example below.



The "Image is a clickable Link"—evasion tactic.

The DocuSign impersonation—the "Display Name Stuffing" of the sender's long email address.

This is a real-world example of a phishing email reported in the KnowBe4 Defend platform, part of a campaign by the threat actor group **Tektite Vulture**. This phish demonstrates the use of both the "Display Name Stuffing" deception tactic and the "Image is a clickable Link" evasion tactic.

Threat Group Name: Sulphur Kestrel

Primary Motivation: Financially motivated, specializing in high-fidelity Microsoft 365 Credential Harvesting and session-based Multi-Factor Authentication (MFA) bypass.

Origin: Indonesia/Singapore

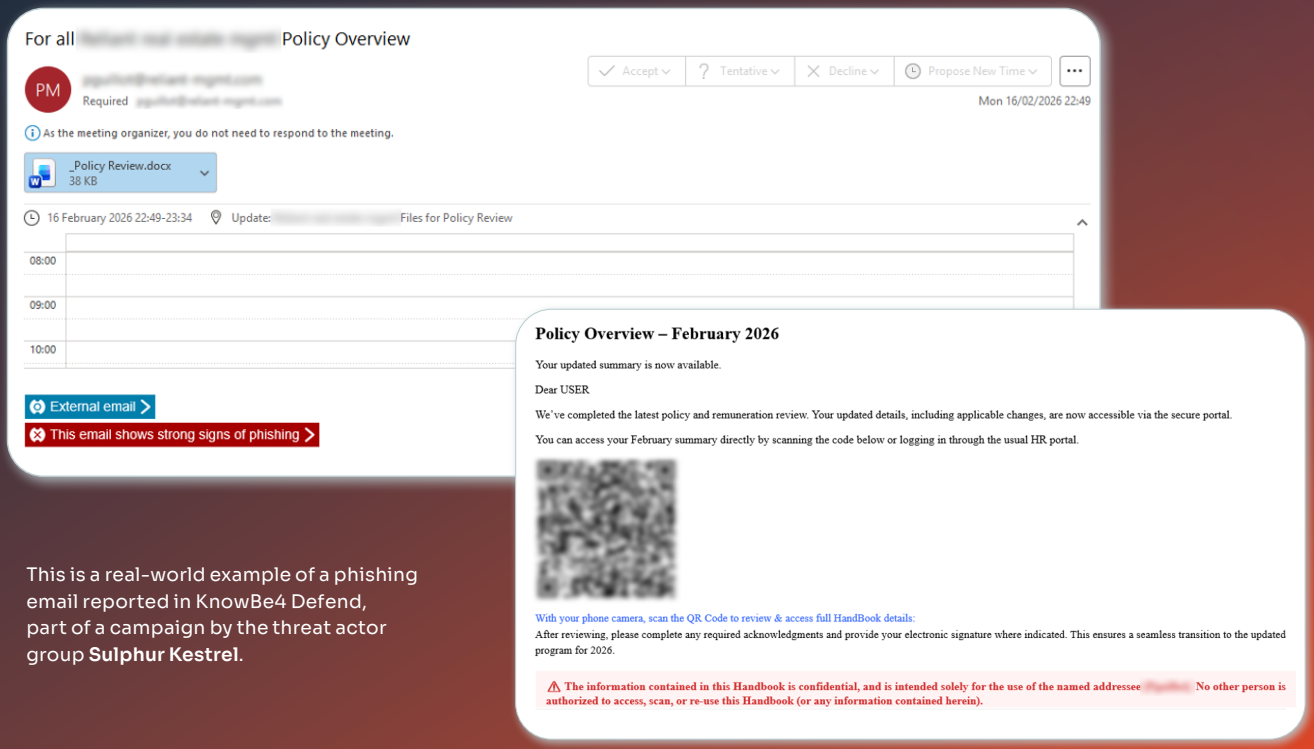
Targets: Indiscriminate global targeting of the private sector, specifically focusing on enterprises using Microsoft 365 and high-reputation corporate VPN environments.

Volume and Tempo: Sulphur Kestrel utilizes a "Just-in-Time" infrastructure, intentionally aging batched domains for weeks to evade "Newly Registered" security filters.

Key Tooling: Expert utilization of the Greatness or NakedPages Phishing-as-a-Service (PhaaS) kits. These tools are specifically designed for reverse proxy operations, allowing the attacker to act as a live bridge between the victim and the legitimate login portal.

Signature Tactics, Techniques, and Procedures (TTPs):

- **Social Engineering Hook:** Utilizes "Policy Review" and "Action Required" lures. Frequently employs Calendar Invite Injection (.ics), which forces email clients to auto-process the message as a meeting, triggering direct system notifications and bypassing traditional inbox spam filters.
- **Evasion:** This threat evades detection by hiding backend servers through geofencing, JavaScript obfuscation, and Cloudflare tunneling. It further bypasses static analysis using split/reverse obfuscation, fragmenting malicious code to prevent tools from flagging keywords like "password" or "login."
- **Deception:** Using Reverse Proxy (AiTM) architecture, this threat mirrors Microsoft login pages with perfect visual accuracy. To evade detection, it employs VPN Provider Redirects—analyzing visitor IPs to reroute security vendors and corporate VPNs to benign sites like Google.



This is a real-world example of a phishing email reported in KnowBe4 Defend, part of a campaign by the threat actor group **Sulphur Kestrel**.

Conclusion

The data reveals a threat landscape defined by automation, scale, and evasion. As adversaries deploy sophisticated, multi-day campaigns across disposable infrastructure, relying on easily manipulated indicators to secure the organization is a losing battle. The ultimate lesson drawn from these APT profiles is that reactive filtering is no longer sufficient.

To effectively break the cycle and anticipate the next strike, organizations must anchor their defenses to the only metrics attackers don't change: their core behaviors, operational tactics, and fundamental methodologies.

Teams Phishing: The Inbox is No Longer the Only Frontline

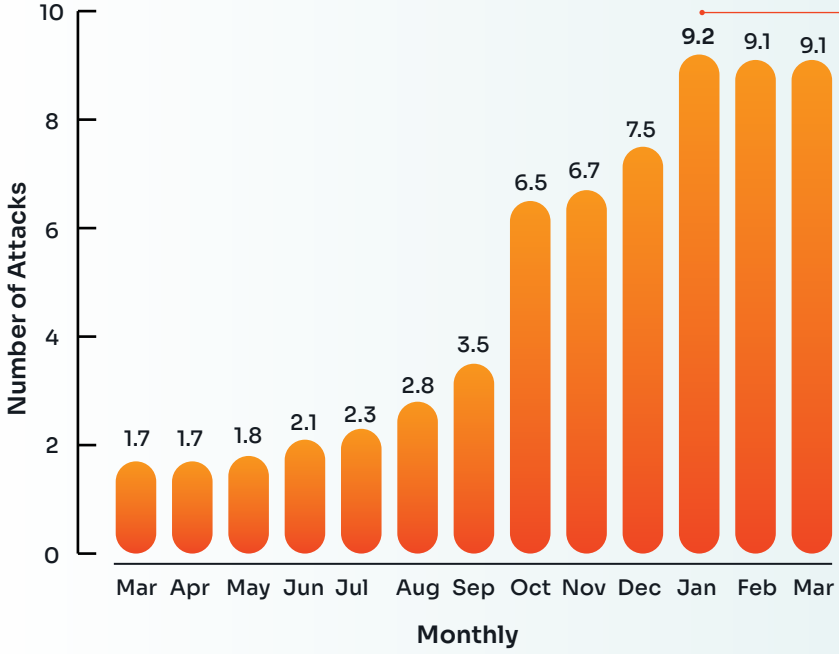
While we've spent years training our eyes to spot phishing emails, the threat landscape is shifting as social engineering evolves beyond the inbox. Today's attackers are increasingly favoring multi-channel campaigns, a trend already proven effective in smishing and phone-based attacks, to bypass traditional defenses and catch users off guard. This strategic pivot toward cross-platform innovation is exposing a critical, often under-protected vulnerability within our organizations: Microsoft Teams.

As Microsoft Teams becomes the central nervous system for global collaboration within organizations, threat actors are following the action to exploit a unique trust gap. Unlike the formal pace of email, Teams is built for speed and informality. This high-speed mindset often causes users to prioritize a quick reply and engagement over a careful security check.

Attackers are banking on this perceived safety, turning our primary collaboration tool into their path of least resistance. While the primary attack vector remains email, transitioning an attack to Teams allows threat actors to extend the kill chain and create more avenues for success. These threats are particularly dangerous as Teams allows an attacker to communicate consistently with a victim over multiple messages. This enables them to build a rapport and a sense of legitimacy that is much harder to achieve through traditional channels.

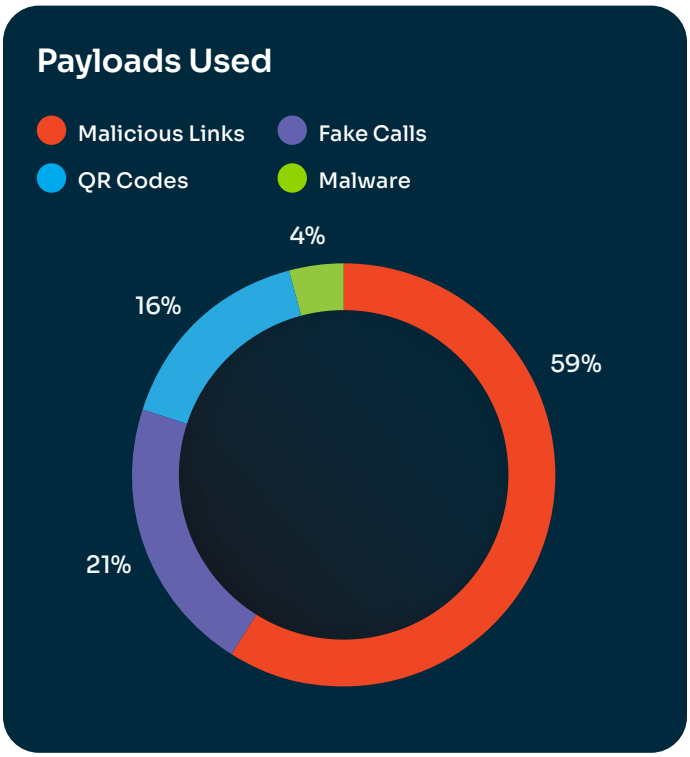
The data tells a clear story: as hybrid work cements itself globally, our Threat Research team has tracked a **41% surge in Teams-based attacks** over the last six months (October 2025 – March 2026), highlighting a critical surface area of risk.

Monthly Average Number of Teams Attacks Reported

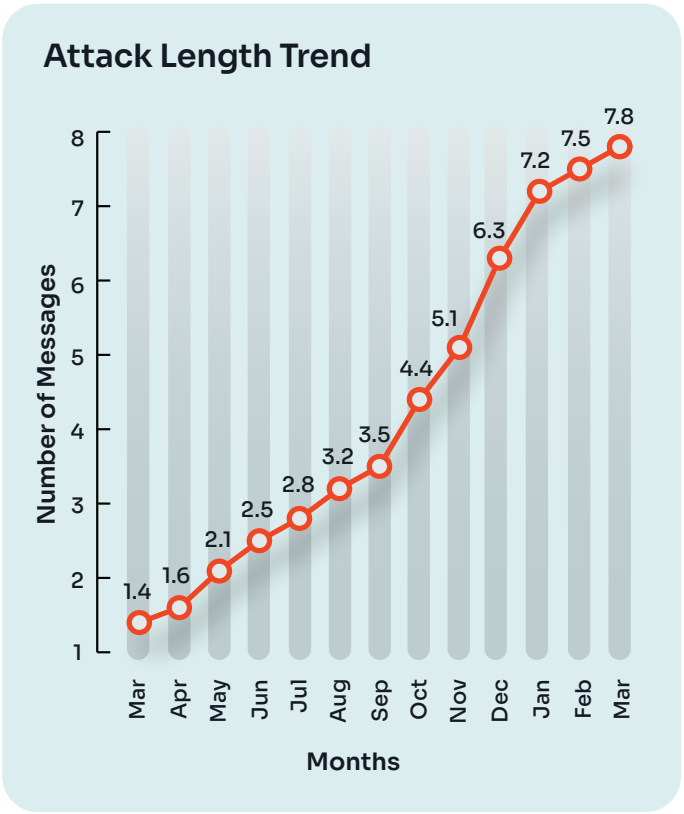


Attackers follow every available avenue. While still emerging, Microsoft Teams attacks are surging, peaking in **January 2026**. This rise exploits the "Chat with Anyone" feature—enabled by default—which lets users initiate chats with any email address, regardless of whether the recipient has a Teams account.

Social Engineering and Payloads



Deepfake Teams calls are an emerging threat, currently representing just under 5% of all call-based attacks. This attack can be conducted through both video and audio, with our analysts identifying that static audio clips are still representing the majority of these attacks at 65%, and are primarily utilized for VIP impersonation attacks. We are starting to see more deepfakes occur during pseudo-live calls. In this scenario, attackers use dynamic, real-time interactions to pressure victims, leaving them little time to question the legitimacy of the encounter before being manipulated into action.



The average volume of messages per attack campaign, though small, has **increased by over 450%**, rising from one to nearly eight messages!

This shift shows that attackers are prioritizing sophisticated social engineering and personalization in their attacks. This technique allows the attacks to establish contact and maintain a rapport before delivering their malicious payload. This heightened level of sophistication allows the attacker to bolster their credibility, significantly increasing the probability of a successful compromise.

Social Engineering and Payloads

Social Engineering Techniques

- Mirrors normal Teams communication (Chat/Pleasantries)
- Impersonation (targeting IT, HR, CEO, Finance)
- Creation of Urgency (setting deadlines for negative consequences)
- Heavy use of social engineering (due to nature of Teams)
- Some attacks are a single email attack, whereas others feature longer campaigns

Most Common Impersonations in Teams (2025 - 2026)

Cybercriminals target roles that carry authority or technical necessity to ensure high-pressure compliance.

1. Information Technology
2. HR (both HR employees and HR platforms like Workday)
3. CEO Impersonation
4. Finance

6-Month Timeline Representing Most Impersonations in Teams Attacks

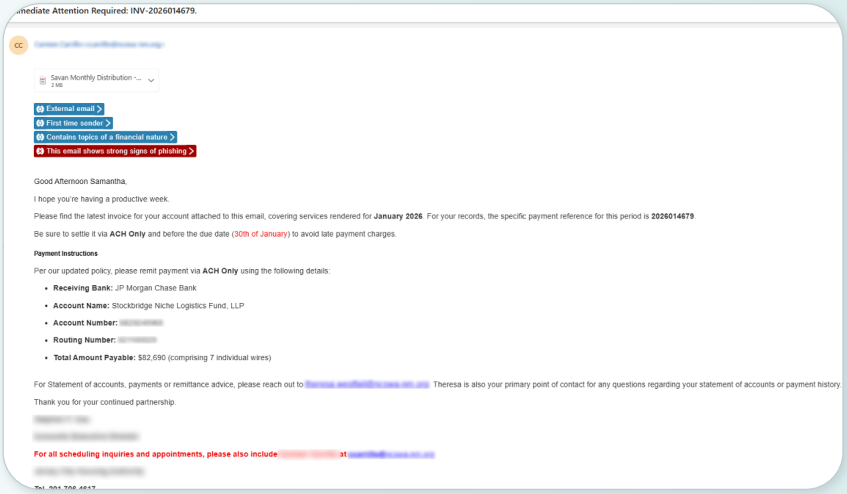
Almost every month is different! This represents how this emerging threat is quickly pivoting depending on attacker priority and responding to time of year.

The Anatomy of a Multi-Channel Attack

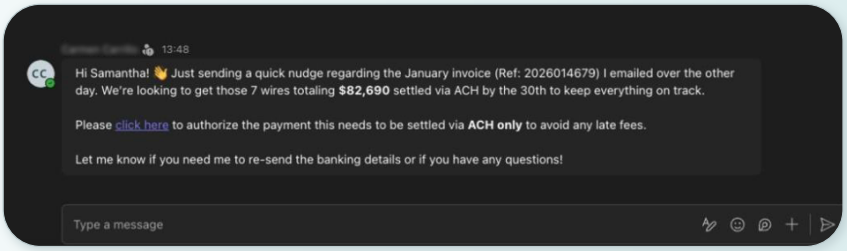
A defining trend is the shift from single-vector attacks to multi-channel orchestration. As email defenses improve, attackers are using Microsoft Teams alongside email to manufacture credibility. By initiating a request via email and "following up" via a Teams message, they exploit the platform's informal nature to validate their identity across different environments.

This cross-platform movement significantly expands the attack surface. When a request appears in both an inbox and a chat thread, it creates a false sense of legitimacy that bypasses traditional mental filters and seamlessly integrates the exploit into the professional workflow.

Step 1 **Initial Email:** Victim receives a phishing email asking for something (ex. Payroll update, urgent IT ticket).



Step 2 **Reinforcement on Teams:** Minutes later, a direct message from the same impersonated entity urging the recipient to complete the action or resending the payload.



Nearly 1 in 5 (17.38%) of all Microsoft Teams attacks are now multi-channel. These attacks originate in the inbox to set the stage but migrate to Teams to deliver the final payload.

Why Multi-Channel Attacks Work

1. Context Switching:
Brain resets security filter from email to Teams

2. Verification Illusion:
Feeling of verification from seeing the sender on two separate platforms

3. Bypassing Technical Scrutiny:
Attacker bypasses email protection by sending the payload directly on Teams

Case Study 1: IT Impersonation Leading to a Call and Remote Access

This case study demonstrates an impersonation attack that pivots from a simple chat message to a real-time, deepfake-assisted voice call to gain remote control over a victim's machine.

Tactical Methodology: From Social Engineering to Account Takeover

The attack lifecycle is characterized by four distinct phases:

- **Establishment of Trust:** Attackers impersonate internal IT personnel, utilizing sophisticated social engineering to bypass initial skepticism.
- **Psychological Pressure:** The use of manufactured urgency and negative consequences compels the recipient toward immediate, reactive behavior.
- **Channel Pivot:** The transition from text-based messaging to a live call serves as the campaign's most sophisticated element, significantly limiting the victim's "think time."
- **AI-Enhanced Exploitation:** On the call, the attacker employs a dynamic deepfake voice, allowing the attacker to respond to queries in real-time. By getting the victim onto a call, the attacker secures remote access to intercept MFA credentials, resulting in a full Microsoft account compromise.

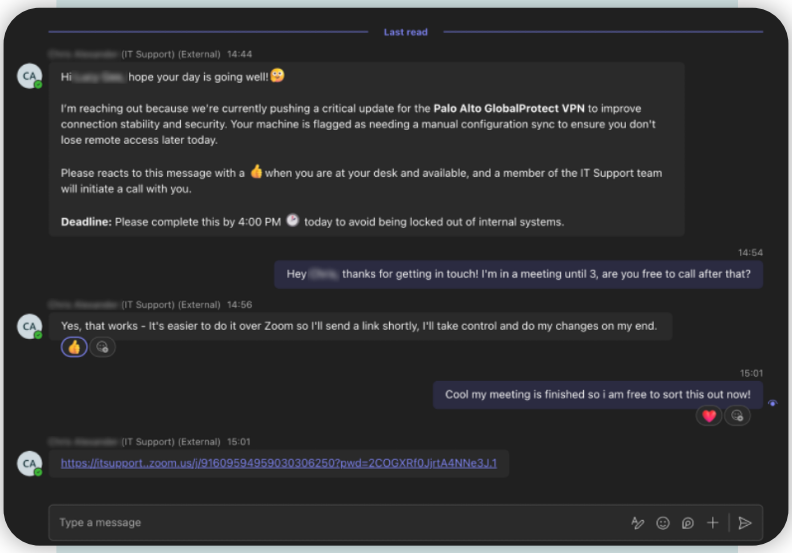
As this is largely an emerging threat, it is worth noting that Microsoft Teams attacks have the same potential consequences for the victim as phishing emails:

- Download of malware onto the computer
- The attacker connecting the victim's Microsoft account to their own device
- Unauthorized access to files and sensitive data
- Attempting to disable security measures for long-term persistence
- Checking the victim's browser for saved passwords
- Deployment of ransomware
- Lateral movement through company network

Once the Teams account is compromised, the attacker can lay dormant for as long as they want before moving laterally through the company network, exfiltrating sensitive data and gathering intelligence to fuel further targeted attacks.

Steps an Attacker Takes to Gain Remote Access And Intercept the MFA Code

Teams Chat (Urgency)
Coerce a user onto a call by fabricating an emergency in a Teams chat.



Zoom Call
Establish trust and rapport by moving the conversation to a live video call.

Screen Share
Convince the victim to share their screen to gain visual access for troubleshooting.

Request for Remote Control
Gain full operational control by requesting and being granted remote desktop access.

MFA Intercept
Intercept the critical MFA code by viewing it on the shared screen.

Case Study 2: HR Platform Impersonation

This next case is a link-based credential harvesting attack that exploits user trust in a critical, high-reputation internal system.

Key Tactics of this Attack:

Platform Impersonation: The attacker impersonates a major HR platform — Workday. This is highly effective because attackers are leveraging a widely known platform used by organizations, leading employees to apply less scrutiny.

Personalization and Legitimacy: The message is heavily personalized with the Workday logo and a changed display name and follows the template of a legitimate Workday notification.

Contextual Timing: The attack is timed for a specific part of the year to target a data review, exploiting a realistic corporate workflow to increase credibility.

Link-Based Payload: These attacks typically rely on a link that redirects the user to a perfect mirror of the login page to steal their credentials.

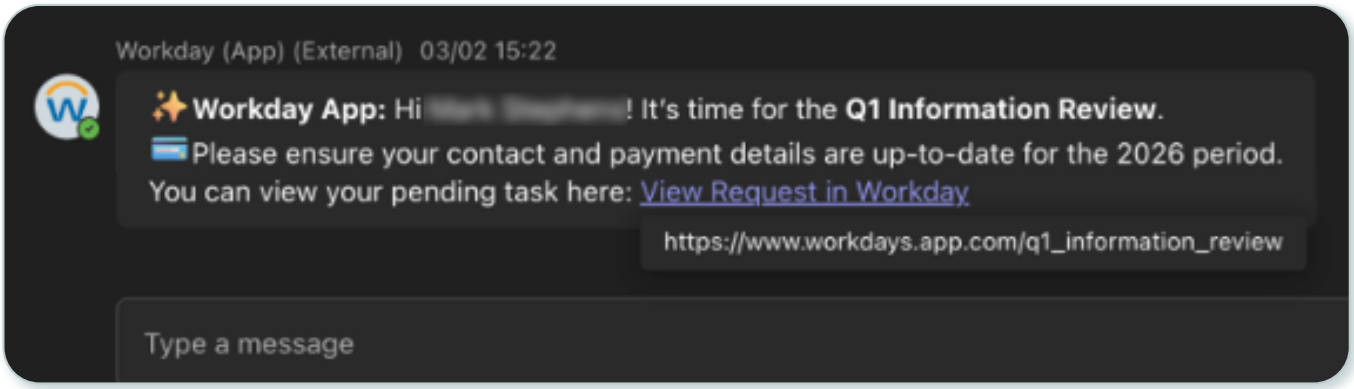
Potential Consequences for the Victim:

Credential Theft: Acquiring the victim's actual Workday account credentials, which grant access to a wealth of personal data, including phone numbers, home addresses, and payroll information.

Lateral Movement: Obtaining a reusable username and password that can be tried in other company applications or systems (credential reuse).

Flexible Payload: The malicious link is versatile and could be configured to redirect the user to a variety of other credential harvesting pages or to download different malicious files.

Example Below:



With Microsoft Teams and other collaboration software as a new vector, attackers are weaponizing the inherent trust of the platform against users. Adding collaboration software to attackers' multi-channel strategy significantly expands the available attack surface, turning a collaborative safe space into a vulnerability. To reduce the risk that this attack vector poses to organizations, collaboration attacks need to be treated with the respect they deserve, with human risk management at the forefront. Though small in volume currently, this emerging threat vector has already been shown to be profitable for attackers.

Adversary-in-the-Middle (AiTM): The Stealthy New Standard

Credential harvesting remains the primary payload of choice for modern adversaries. Currently, 60.13% of all identified attacks rely exclusively on malicious links, while 90% of malicious attachments incorporate embedded credential harvesting pages. Threat actors continue to invest heavily in this methodology, increasing technical sophistication to eliminate traditional indicators of compromise and bypass user detection.

Adversary-in-the-Middle (AiTM) phishing is a technique that uses dedicated tooling to establish a proxy between a target user and a legitimate login portal for an application.

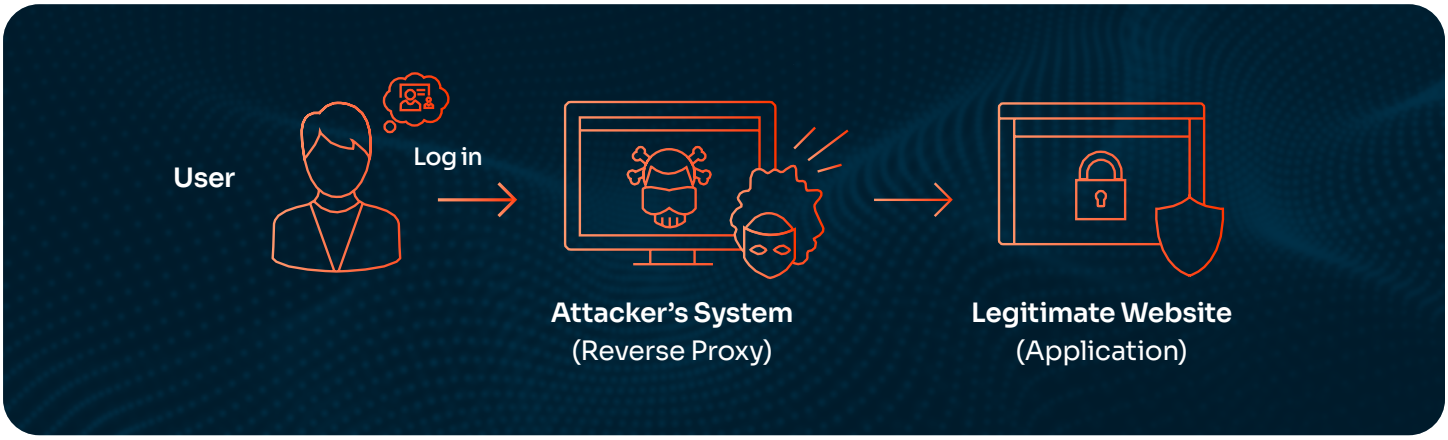
Imagine pulling up to a five-star hotel. A valet in a crisp uniform takes your keys and gives you a ticket. You watch him drive your car into the legitimate hotel garage. What you don't see is that the valet isn't a hotel employee; he's a thief. He still parked your car in the right spot, but he just made a 3D mold of your key and a copy of your registration while he was in the driver's seat.

The Attack's Power Comes From Intercepting the User's Data Without Altering Their Experience

The deployment of a malicious proxy ensures that the spoofed login interface appears identical to the authentic site. This is because the target is, in fact, logging into the legitimate site, but their connection is routed through an attacker-controlled intermediary.

This approach makes AiTM attacks uniquely effective for three primary reasons. First, the method maintains total authenticity. Because the login page is a live proxy of the actual service, the interface is indistinguishable from the real thing, which makes the compromise nearly invisible to the user. Second, it enables silent data interception. By positioning their system as an intermediary, the attacker observes every interaction in real-time to harvest credentials and capture the active session ticket. Finally, this leads to an immediate account takeover. Armed with a stolen session ticket, the attacker can bypass MFA and seize full control of the account without the user ever realizing they were targeted.

Adversary-in-the-Middle (AiTM) Process



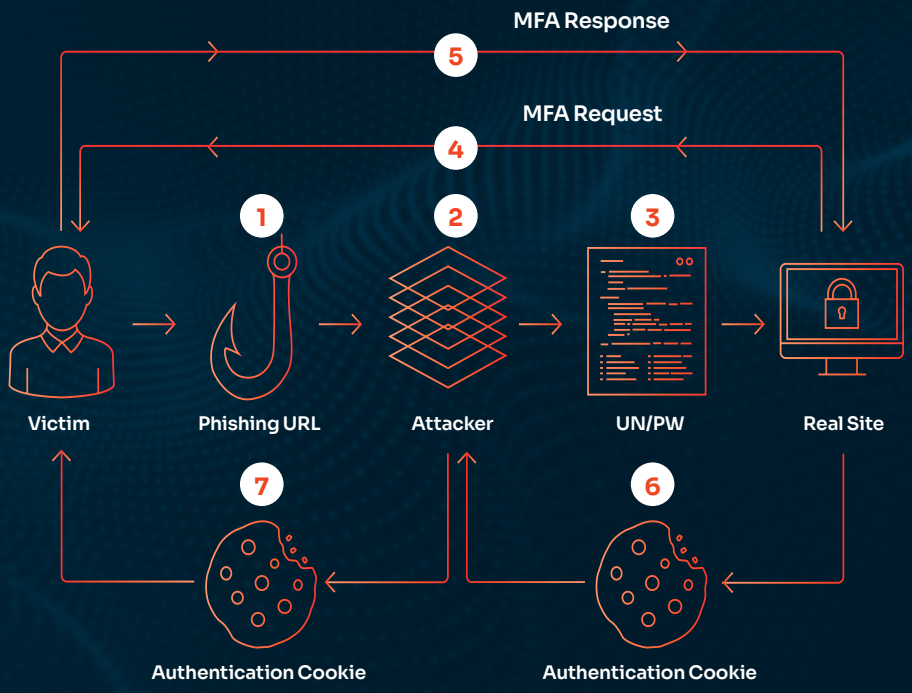
Reverse Web Proxy

The malicious proxy is the core component enabling the sophisticated nature of AiTM attacks. This technique sets up a live proxy between the victim and a legitimate login portal, allowing the attacker to observe all interactions and steal credentials and the authenticated session cookie. Of the various proxy methods available, the reverse web proxy is arguably the most scalable and reliable approach from an attacker's point of view. This method is highly effective because the victim is interacting with a real site, but instead it directly allows the attacker to bypass traditional security measures like MFA.

The technique is also accessible due to the availability of open-source tooling, which automates the process and is leveraged by a wide range of threat actors. Popular tools demonstrating this method include Evilginx, Modlishka, and Muraena. For instance, the threat group **Sulphur Kestrel** expertly leverages this AiTM architecture, utilizing PhaaS kits like Greatness or NakedPages for their reverse proxy operations.

MFA Bypass Attack Process

1. Victim sends US/PW
2. Attack capture US/PW as it passes through reverse proxy
3. Attack relays US/PW to legitimate site
4. **MFA Request:** MFA request is sent to victim by legitimate site
5. **MFA Response:** MFA request is approved by victim
6. Legitimate site sends authentication cookie back to attacker's reverse proxy
7. Attacker sends authentication cookie backto victim



Top Phishing-as-a-Service (PhaaS) Toolkits

- Evilginx
- EvilnoVNC
- Kratos
- Modlishka
- Muraena
- Quantum Route Redirect
- Starkiller

Ask The Expert

James Dyer
Head of Threat Intelligence

Phishing-as-a-Service (PhaaS) Toolkits

- These toolkits automate the entire attack lifecycle, from deploying high-fidelity brand clones to real-time data exfiltration via the Telegram Bot API. The Kratos toolkit specifically has been noted for evading bots, where the real malicious payload only activates upon human interaction.

The Impact, the Motivation and the Defense of Reverse Proxy

THE THREAT

Reverse Proxy Surge

139.22%
Surge from
Sept 2025 – March 2026

Reverse Proxy surge!
This technique is commonly used for attacks, but our threat researchers have spotted a large spike in their dataset where Reverse Proxies are now being used more than ever before.

THE MOTIVATION

Why Attackers Choose AiTM

Ease of Use
Automated tooling like Evilginx

MFA Bypass
Steals active session tickets

Looks Real
Live proxy of the real login page

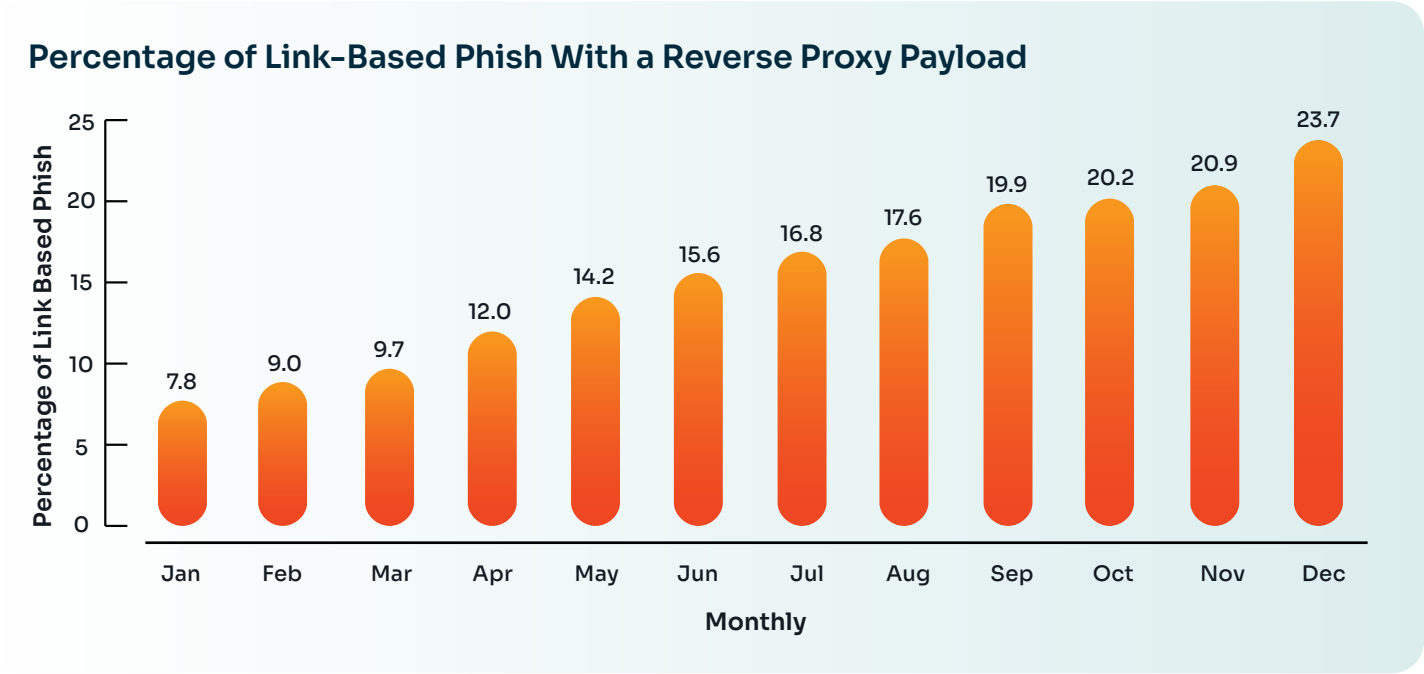
It Works!
High success rate

THE DEFENSE

How to Spot a Reverse Proxy

Domain Verification:
Check the URL. Phishing sites often use similar but slightly altered domains (e.g., micr0soft.com instead of microsoft.com). This is the most crucial user-level check.

Network Traffic Analysis:
Monitor connections. Use tools to analyze incoming and outgoing traffic for signs of a man-in-the-middle proxy connection.



VPN Detection and Redirection

Attackers often bake additional layers into their attacks to make it convoluted for security professionals to understand the full picture and scope. This has traditionally been achieved by geolocation checks and abusing the single-use link approach, voiding the attack being replicated by an interested third party. However, our threat researchers have found that many attackers are now checking connections originating from top corporate VPN providers as an additional layer. If a connection is identified as coming from one of these VPNs, the attacker's phishing infrastructure will redirect the visitor to a benign site. This technique is used to gather more intelligence on potential victims and wait until the target accesses the phishing site from a potentially less-secure or vulnerable internet connection.

The list below showcases a selection of top corporate VPN providers and the corresponding benign sites to which attackers redirect traffic:

Palo Alto GlobalProtect
tesla.com, forbes.com, quora.com, medium.com, cnn.com, x.com, walmart.com, disneyplus.com, shopify.com, gitlab.com

Checkpoint Secure Remote Access
stackoverflow.com, twitch.tv, soundcloud.com, dailymotion.com, paypal.com, gitbucket.org

Citrix Secure Private Access
hulu.com, nytimes.com, washingtonpost.com

OpenVPN Access Server
bbc.com, github.com, wired.com, box.com

Cisco AnyConnect
drive.google.com, onedrive.live.com, soundcloud.com, slack.com, dropbox.com

The Agentic Shift: Anticipating the Era of AI-Driven Threats

Over the past couple of years, AI has triggered a fundamental shift in how organizations operate, with worldwide spending on AI forecast by Gartner to reach $2.52 trillion in 2026, a **44% year-over-year increase**. We have moved beyond simple chatbots and reactive assistants into the era of agentic orchestration, where autonomous AI agents are woven into the fabric of global workflows to summarize, filter, and execute complex tasks. However, the same efficiency gains benefiting the enterprise are being mirrored by the adversary. Attackers are now leveraging agentic tools to weaponize reconnaissance and scale high-fidelity threats that were previously impossible to automate.

We previously predicted that "in the near future, some form of AI will be used in almost every phishing attack." In the past six months alone, **85.8%** of phishing attacks were AI driven.

Criminal AI Usage

Reconnaissance

AI has industrialized the reconnaissance phase. By organizing massive datasets to reveal hidden patterns, attackers can ingest sensitive information and historical breach data at an unprecedented scale. This automated intelligence gathering facilitates hyper-targeted profiling, where public documents and social media are harvested to build precise maps of an organization's internal projects and supply chain vulnerabilities.

Development

Users are inherently motivated and satisfied by effective personalization; we gravitate toward experiences that feel tailor-made for us. This is evident in our daily lives through the seamless convenience of curated shopping recommendations or the personalized "wrapped" summaries of our digital habits. Cybercriminals understand this psychological pull perfectly and have weaponized it; by mirroring this same level of relevance in their lures, they exploit our natural desire for individual attention to bypass our critical judgment.

Artificial intelligence scales this into vast, automated attacks, pulling data from the reconnaissance stage directly to the phish using the following techniques:

- Addressing recipient by name
- Aligning email to company brand
- Impersonating closest social graph relations
- Timing attack with relevant projects and events

Payload

Polymorphic phishing campaigns consist of a series of almost identical emails, differing only by a few small details.

AI has supercharged polymorphic attacks, moving beyond simple variations in subject lines to the automated generation of entirely unique phishing content for every recipient. This allows attackers to pivot from bulk "spray-and-pray" tactics to deploying thousands of unique and personalized lures simultaneously. With matured AI toolsets, a single campaign can now instantly generate and distribute multiple payload formats to further bypass traditional detection.

Common techniques used in polymorphic, AI-driven phish:

Autogenerated malicious links, often combining URL shorteners, obfuscation techniques, and compromised reputable sites to mask the final destination.

Unique variations of malicious attachments changing the file's signature and content while remaining personalized to the target victim.

AI-generated malware, each remodeled with obfuscated variables and structure to disrupt signature-based detection.

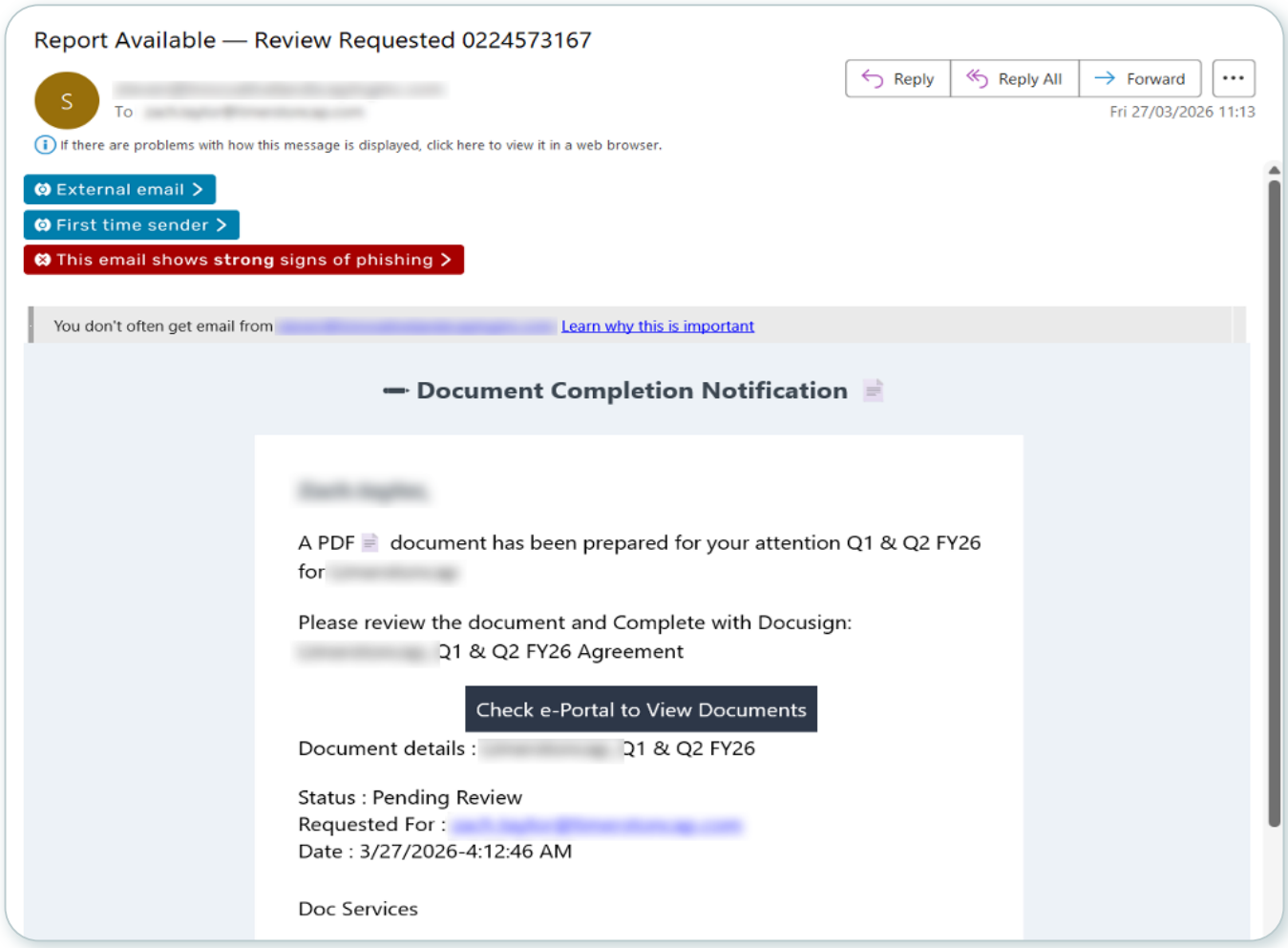
In the past three years, we have seen phish using polymorphic elements rise from **56.9%** in 2024 to **67.3%** so far this year.

Execution

The ultimate success of the campaign hinges on the phish itself. By combining insights from the previous stages, attackers can craft a seemingly legitimate phish, a direct byproduct of social engineering and high levels of personalization.

To take this sophistication to the next level, the attacker can compromise a business domain to send their carefully curated attack. Ensuring that authentication passes checks is the final nail in the coffin to bypass traditional solutions and land their attack in their victims' inboxes.

Polymorphic BEC Phishing Campaign with Elements of Personalization



We estimate that AI-driven attacks are **seven times more efficient** than those relying on manual reconnaissance. By automating the research phase, AI has effectively eliminated the manual labor previously required to hit high-value targets, allowing attackers to scale sophisticated campaigns at an unprecedented pace.

Diversifying the Payload

AI has transformed the phishing lifecycle from end to end, allowing attackers to abandon static templates in favor of dynamic, creative payloads. In our digital-first world, the abundance of legitimate audio and video content, from podcasts and corporate webinars to TikToks, has become a goldmine for exploitation. By harvesting this data to build precise linguistic profiles and training generative models on the results, threat actors can now clone executive voices with startling accuracy.

These clones are increasingly deployed in live, "calibrated" scenarios on platforms like Zoom, where attackers leverage the contextual authority of leadership to manipulate employees. While these tactics sound complicated, tools like ElevenLabs, Resemble AI, and Voice.ai have significantly lowered the barrier to entry, making sophisticated deepfake attacks a mainstream threat.

By eliminating language barriers and syntactic red flags, these tools allow attackers to scale high-fidelity, multi-vector campaigns globally with minimal effort.

MP3 Files in Phishing Emails

38.1%

Increase in the number of MP3 files being used in phishing since the start of the year

73.3%

Increase in malicious MP3 files as a share of total MP3 files between 2022 to 2026

72.9%

Increase in the average size of MP3 files

With an average MP3 file length of two minutes and 13 seconds, attackers need to balance providing enough audio to be convincing while avoiding red flags that could potentially expose them.

Common Words/Phrases Found in MP3 Files

Let's see what the attackers have to say:

Average Size of Attachments Has Increased

We are witnessing year-on-year growth in file sizes, a trend driven by the increased integration of AI in phishing emails. Attackers leverage auto-generated content within the file to overwhelm security systems from scanning the content, effectively obscuring the malicious payload.

142.2 KB	2023
144.8 KB	2024
156.7 KB	2025
183.7 KB	2026

111.8% increase in unique malware signatures since the start of 2026 while the size of these malware samples has **increased by 9.7%.**

How Attackers Are Leveraging AI to Attack Security Vendors Directly

While AI has accelerated the speed and efficiency of phishing, it has also provided cyber criminals with the tools to systematically undermine traditional security perimeters. By disrupting core detection parameters, such as NLP and NLU, these AI-driven threats bypass machine defenses, shifting the entire burden of detection to the human — the final and most vulnerable point of failure.

Common techniques used to bypass security vendors include:

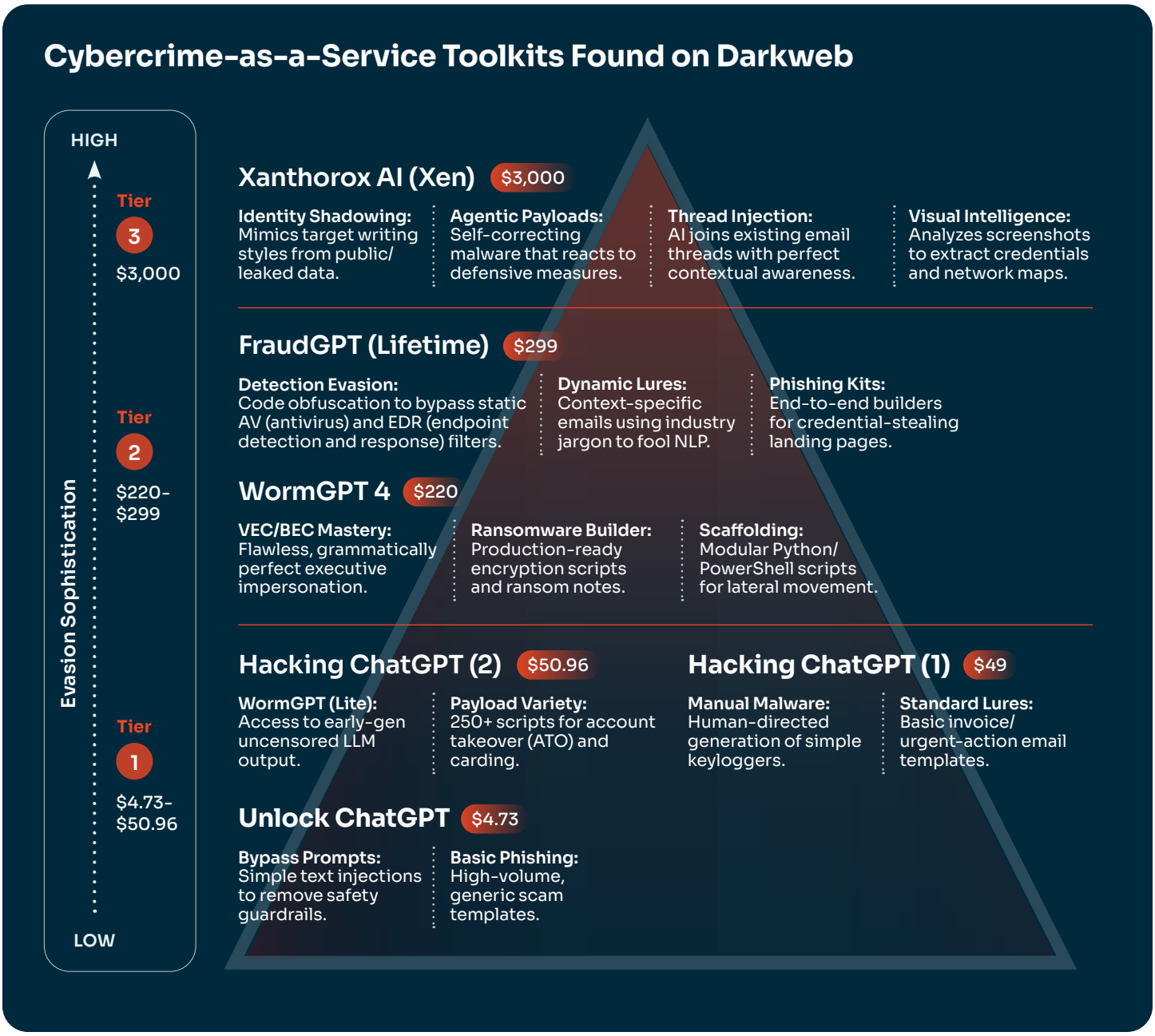
- White-on-white text embedded in documents
- HTML smuggling
- Hidden HTML or markdown instructions in webpage source code
- Zero-pixel images containing malicious metadata
- Microscopic footnotes buried in large documents
- Fake positive signaling to poison data models
- Flood attacks to build trust, preventing anomaly detection and abuses social graphing

Cybercrime-as-a-Service

To get a clear picture on how AI has transformed the threat landscape, our researchers investigated the dark web to identify the resources currently available to attackers and determine how easily they can be acquired.

To allow the attackers to operate at scale for the above attacks, a mini ecosystem has been established, where tools for jumping over LLM safeguards and creating malicious agents are a hot commodity.

These methods are now being packaged into Cybercrime-as-a-Service (CaaS) toolkits, enabling even low-skilled threat actors to execute advanced attacks. The democratization of AI has fueled this thriving CaaS market on underground forums, posing a persistent challenge to AI safety and governance.



Ask The Expert

James Dyer
Head of Threat Intelligence

- **The above table is a stark illustration of the industrialization of AI-driven cybercrime through the Cybercrime-as-a-Service (CaaS) ecosystem.** The most interesting trend is the direct correlation between technical abilities and pricing, which ranges drastically **from $4.73 to $3,000**. Ultimately, the data shows that attackers are scaling beyond simple LLM jailbreaks to deploy integrated, full-spectrum cybercrime campaigns, proving these toolkits are quickly turning AI manipulation into a mainstream commodity.

The Next Step: Prompt Injection

AI innovation shows no sign of slowing, and as it evolves, so does the creativity of the adversary. In fact, we are already seeing a new class of threats: emails engineered to manipulate the AI systems they use. By embedding instructions that are invisible to humans but legible to machines, a technique known as indirect prompt injection, attackers can hijack an AI's logic as it processes the data.

Once subverted, a standard AI assistant transforms into a "malgent" — a malicious agent operating with the trusted permissions of an employee. These agents often have broad access to databases, APIs, and internal comms, they become the ultimate inadvertent insider threat, executing malicious tasks with the authority of a legitimate user but without any human oversight.

In this environment, passive defense is no longer viable; staying ahead of the threat requires a proactive, AI-augmented security posture that matches the speed and sophistication of the modern adversary.

The result is a pivot in the threat landscape: attackers are no longer just socially engineering humans; they are socially engineering AI.

Calendar Invites: Infiltrating the Corporate Schedule

Modern phishing has found a path of least resistance: the calendar invite. By abusing the .ics file, a universally trusted, text-based file — attackers bypass standard security solutions that typically categorize these files as a benign scheduling tool. Unlike traditional phishing, which demands active engagement, this technique leverages an automatic delivery mechanism: inserting malicious events into a user's schedule regardless of whether the initial email is ever opened.

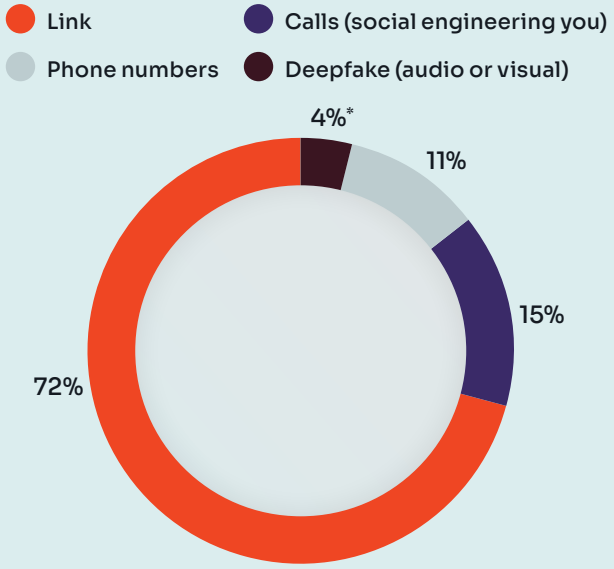
The danger of this attack is rooted in its calculated exploitation of a psychological blind spot. By shifting the field of play from the scrutinized and busy inbox to the relative sanctuary of the corporate calendar, threat actors bypass traditional skepticism. In the fast-paced environment of a modern workday, even the most security-conscious employees operate with an implicit trust in their calendars. For professionals perpetually jumping from meeting to meeting, the calendar is no longer just a tool, it is a to-do list for their day. This trust is further weaponized when a system-generated notification triggers a reminder for the event, catching the user in a reactive, high-speed mindset. Users are far more likely to just join the meeting than verify its origin. This reflexive engagement allows the attacker to bypass a user's natural suspicion, transforming a routine calendar alert into a high-consequence entry point for credential harvesting. By merging technical obfuscation with this "default-to-trust" behavior, attackers effectively neutralize both automated filters and human intuition in a single, silent stroke.

Calendar Invite Phishing has **surged by 49%** in the last six months, confirming this method's escalating threat.

Common Tactics

- Sense of urgency
- Financial linguistics (e.g., salary discussion)
- A phone number (typically a fake support line)
- Impersonation (used as a social engineering technique to increase the perceived legitimacy)
- Predominantly link based

Calendar Phish Payload Breakdown

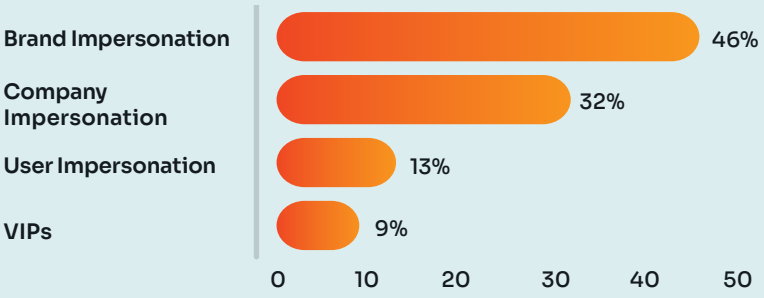


*Deepfakes are starting to mature and become more mainstream. Threat researchers expect this to increase over the next six months.

The Power of Impersonation

A **staggering 85% of calendar phish** use impersonation tactics to manufacture legitimacy.

Impersonation Types

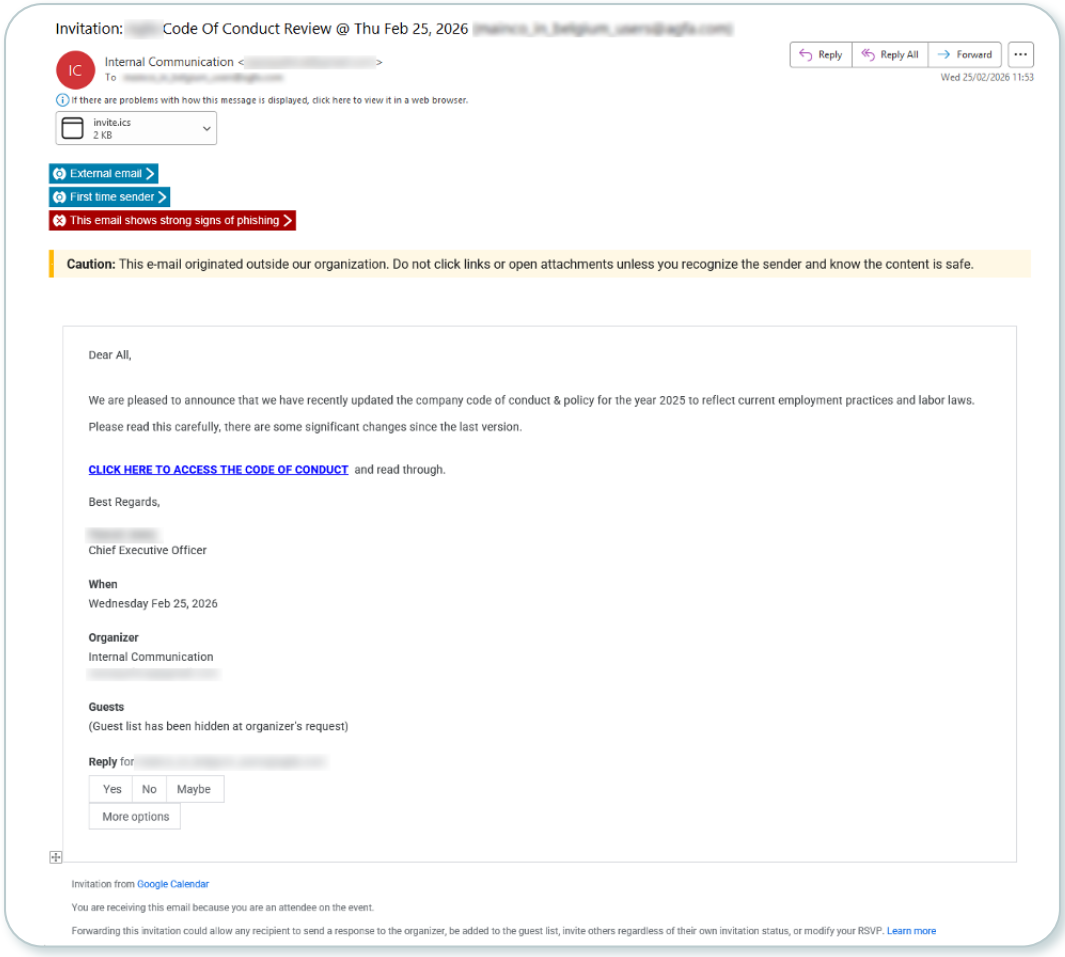


Approximately half (46%) of impersonations are from well-known, common brands, including:

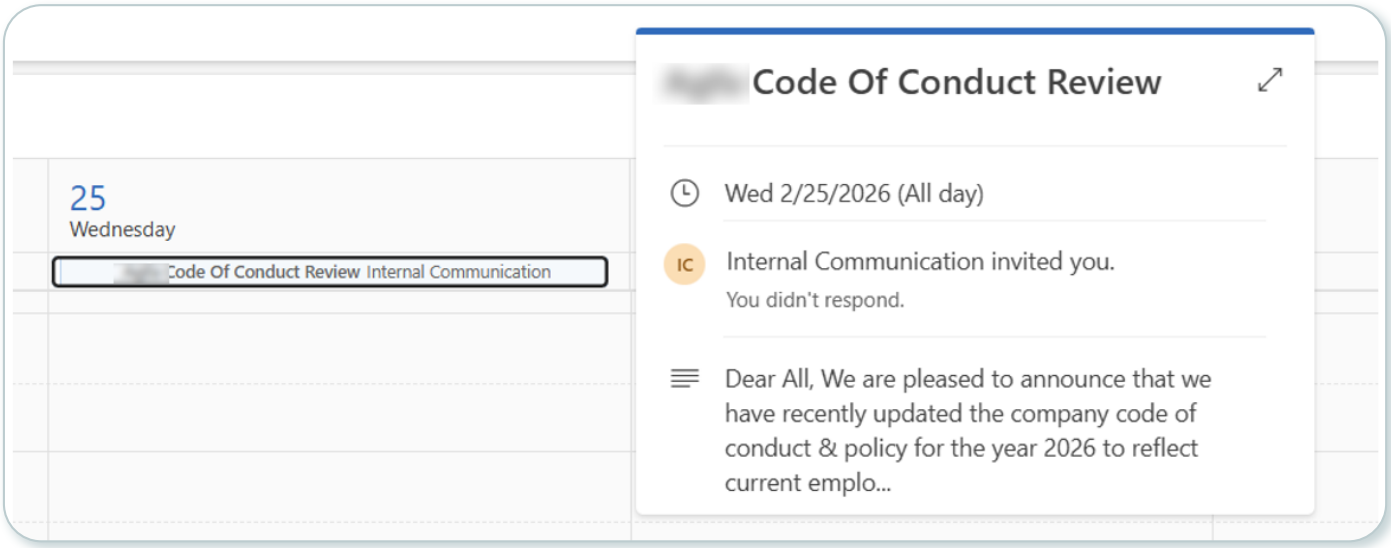
PayPal Microsoft Google zoom bitcoin

Beyond malicious links, social engineering, and deep fakes, these attacks serve a scouting function. Attackers are often notified if an invite is accepted or declined, providing them with confirmation that an email address is active and identifying high-value targets for future engagement. By placing both an email in a target's inbox and a meeting on their calendar, this phishing technique effectively doubles the probability of a successful compromise.

Example Below:



How This Looks in a Calendar:



Technical Analysis: Infrastructure Abuse & Reputation Laundering

- Primary Delivery: The attacker abuses legitimate Google Calendar notifications to ensure DMARC compliance. This allows the malicious invite to bypass perimeter defenses under the guise of an internal compliance mandate.
- Redirect Orchestration: The payload is obscured within a multi-stage redirect chain. It leverages a high-reputation Google tracking URL before pivoting to a compromised Ukrainian (.edu) domain.
- Evasion & Exploitation: By utilizing an open redirect vulnerability on the university site, the attacker exploits the link's reputation. This technique ensures the final credential-harvesting destination is hidden from reputation-based detection during the initial delivery phase.

Conclusion

While using a calendar invite may seem like a minor shift in payload, it introduces a sophisticated social engineering twist by infiltrating an environment typically perceived as safe and structured. This technique creates a **double-vector attack**: the initial email delivery and the persistent calendar event.

The surge in calendar phishing is driven by a simple reality: it works. Traditional defenses often classify meeting requests as benign, allowing attackers to exploit a specific cognitive bias. Because users are conditioned to trust their schedules, they are significantly more likely to engage with a calendar notification than a message in a cluttered inbox. Attackers are banking on "calendar fatigue," relying on busy employees who treat their schedule as an absolute source of truth and rarely verify an event's legitimacy before clicking.

2026 Intelligence Brief: Ask the Experts

Through our frequent engagement with readers, we have gained valuable direct insight into the specific organizational challenges you face. This section is designed to address the most common inquiries we receive while providing updates on critical shifts in the threat landscape. Our goal is to deliver actionable intelligence that empowers your organization to better protect its personnel, customers, data and infrastructure.

Q **Our current secure email gateway (SEG) seems to be missing more threats than usual. Are SEGs becoming obsolete?**

A They certainly aren't obsolete; however, they are struggling to keep up with the rising sophistication of phishing attacks. We have witnessed a **31.4%** increase in phishing attacks that are successfully evading SEGs. Attackers are specifically designing more technical attacks, with **61.2%** of phishing emails coming from **compromised business accounts** and **64%** of attacks now utilizing obfuscation techniques, such as invisible characters or links obscured within images to specifically disrupt the detection logic of these gateways.

Q **Are attackers really using artificial intelligence within their phishing attacks? If yes, how much of a problem is it?**

A Yes, in the past six months, our data suggests that **85.76%** of attacks are using artificial intelligence in some form within their phishing campaigns. This has increased year on year from **79.9% in 2024** to **84% in 2025**. AI has lowered the barrier of entry into cybercrime, making it significantly easier to create high-quality, heavily personalized phishing emails at a massive scale while simultaneously extending target surface area.

Q **What type of payloads should my team be concerned about? Where should we be focusing our energy, time, and resources?**

A While every organization faces a different threat landscape, in general, malicious hyperlinks remain a primary vector for attackers, with **60.13%** of phishing attempts containing a link. This payload mechanism offers the versatility to deploy thousands of credential-harvesting sites, implement CAPTCHAs to evade automated scanning, and extract extensive data beyond basic credentials as demonstrated in AiTM.

Q **Are executable attachments a problem anymore?**

A Today **67%** of attachment-based payloads are in the form of PDFs, SVGs, and DOCX files, with executables such as Exe or HTML tumbling down the leaderboard. Researchers have witnessed attackers pivoting to attachments more commonly used daily in business communications in order to fly under the radar. Notably, **90% of these attachments are credential harvesting links**, not direct malware. A growing theme shows attackers gaining an initial foothold into a business and then pushing malicious code organization-wide.

Q **What is one payload you think security professionals should be aware of and have security measures in place to combat?**

A Malicious phone numbers have burst onto the scene as an indirect payload to socially engineer victims outside of email communication. This equates to **7.21% of attacks in 2026** so far to contain a phone number. Common locations of this malicious content include within the subject, in the first line of the email, or within the attachment for extra layers of obfuscation.

Q **Are humans writing these malicious emails or is it an LLM now?**

A The answer, ultimately, is that it's a blended approach. The average length of a phishing email has nearly **doubled since 2022**, jumping from **497 to 1,011 characters**. Attackers are using AI to write longer, more convincing narratives that build trust, making the social engineering aspect much more potent.

Q **We're seeing a decrease in standard QR code attacks. Does this mean quishing was just a fad?**

A While overall QR usage dipped to **3.79%**, the sophistication and obfuscation skyrocketed. As attackers have moved away from "standard" QR codes within the body of the email, they have transitioned into inserting **64.51% of QR codes within the attachment**. This effectively "nests" the payload to obscure it from detection by SEGs, which scan the body but will miss a QR code tucked inside a PDF or SVG file.

Q Who are the most targeted job roles and industries?

A Threat actors consistently prioritize individuals with the keys to the kingdom, which explains why senior executives are the most targeted.

The top five most targeted job roles are CEO, CPO, CFO, VP of Finance, and COO. After the top-10 most attacked job roles, there is a wider range of attack targets — from interns to managers — as it becomes a conversation around individual risk profiles rather than access.

Q How much time do we actually have to train a new employee before they are targeted by a real phishing attack?

A You have to be on it from the get-go! Our data indicates that the average new starter receives their first phish in the **first month of employment**. This suggests that attackers are actively monitoring professional networks and "new joiner" announcements in real-time to identify fresh targets who may not yet be fully integrated into the company's security culture or familiar with standard internal processes.

Q If we block all "new" and "untrusted" domains, will that significantly reduce the number of phishing emails reaching our users?

A Unfortunately, the data shows that blocking "bad" domains is only a small piece of the puzzle. Currently, **61.2%** of phishing emails that successfully bypass security gateways come from **compromised accounts**, with **11.4%** of these compromised accounts coming from trusted accounts within your supply chain!

More than one-quarter **(27.6%)** come from Webmail providers like Gmail or Outlook, and only **13.6%** of phish are sent from dedicated "phishing domains." This means **84.4% of all successful phish now pass DMARC**, rendering traditional "identity verification" defenses nearly obsolete.

Q We've seen a massive increase in phishing emails that appear to come from trusted services like Google, Microsoft, and SharePoint. Is this just a coincidence, or is it a specific tactic?

A It is a very deliberate and highly effective tactic. Our latest data indicates that **22% of all phishing attacks are now sent through a legitimate platform**. By using trusted platforms, attackers are no longer just impersonating these brands; they're abusing the actual infrastructure of these trusted providers. This allows their email to bypass standard domain-reputation filters, as all the authentication results marry up to the legitimate platform and the email lands in your inbox!

Wrapping Up: Staying One Step Ahead in a Shifting Landscape

We hope this edition of the Phishing Threat Trends Report has been an eye-opener. Our goal is simple: to keep you in the know so you can build a robust defense that protects your people, your data, and your brand.

If this year's trends have shown us anything, it's that attackers are getting bolder and incredibly creative with how they land their malicious content in an inbox or an app.

By shifting our tracking methodology within email, we have shown that attacker attribution is the catalyst for transitioning from a reactive defense to a predictive posture. We focused our efforts in this report to showcase how the inbox is no longer the only frontline, and your corporate calendar is being infiltrated to exploit the frantic pace of the digital workday.

Critically, it's not only your humans being targeted by malicious actors but a shift into exploiting your agents and tooling. Adversary-in-the-Middle (AiTM) techniques have dismantled the myth of MFA as a "silver bullet," proving that even our strongest safeguards require constant evolution. As we look ahead, the security of AI emerges as our next frontier, where protecting the integrity of our models is now just as vital as defending the users who interact with them.

Future-Proofing: The Security Revolution

Relying solely on native security or legacy gateways is no longer a strategy — it's a gamble. As cybercriminals pivot and equip their armory with sophisticated AiTM payload and multi-platform attacks, your security stack needs to be as agile and adaptive as the adversaries themselves.

By moving toward a holistic ecosystem fueled by deep behavioral analytics and real-time threat intelligence, your employees become a line of defense. More than a filter, it is continuous, automated coaching that evolves as fast as the attackers do.

Ready to level up your defense? The KnowBe4 team is here to help you turn these insights into action. Let's keep the conversation going and ensure your organization stays ahead of the curve.

Get In Touch

Contributors

James Dyer
Head of Threat Intelligence

James spearheads the Threat Intelligence team at KnowBe4, spending his days uncovering the latest phishing threat trends, understanding emerging methodologies, and analyzing the TTPs of the Crime-as-a-Service ecosystem.

Lucy Gee
Cyber Security Threat Researcher

Lucy is passionate about the intersection of psychology and cybersecurity and the use of behavioral insights to enable people to live and work securely. At KnowBe4, Lucy analyzes the latest phishing campaigns and communicates emerging trends to business stakeholders.

Cameron Sweeney
Cyber Security Threat Researcher

Cameron specializes in understanding the technical aspects of cyberattacks. As a member of the KnowBe4 team, he reverse engineers phishing attacks and malware to identify emerging threats, using statistical analysis to track the evolving threat landscape.

Louis Tiley
Cyber Security Threat Researcher

Louis researches diverse attack vectors, social engineering tactics, and emerging threats. At KnowBe4, he analyzes phishing campaign methodologies and builds tools to automate threat intelligence gathering to identify industry trends and shape cybersecurity messaging.

Jack Chapman
SVP Threat Intelligence

Jack leverages deep insights of the cyber-threat landscape and his extensive R&D skillset to oversee threat research and AI development for KnowBe4 Defend to stop the advanced phishing attacks that defate traditional security solutions. Jack maintains close ties with the global cyber community, particularly the UK's intelligence and cyber agency GCHQ.

Dr. Martin J. Krämer
CISO Advisor

Martin is a CISO Advisor at KnowBe4. He has more than 10 years of research and industry experience in cybersecurity with a focus on human-centered computing. Martin held roles in innovation, research, and technology consulting. He has worked with both public and private organizations on information security and data protection.

About KnowBe4 Defend

An integrated cloud email security solution, Defend delivers AI-powered behavioral-based detection to eliminate the attacks that get through native security and secure email gateways. Leveraging zero-trust and pre-generative models, Defend provides the highest efficacy of detection against advanced threats, including zero-day and emerging attacks, phishing emails sent from compromised accounts, and social engineering. Using dynamic banners applied to neutralized threats, Defend provides real-time teachable moments that continually "nudge" employees into good security behaviors to tangibly reduce risk and augment security awareness.

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 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