CHECK POINT™

cp<r>
CHECK POINT RESEARCH

CYBER SECURITY REPORT 2026

14th ANNUAL EDITION

TABLE OF CONTENTS:

01 INTRODUCTION

02 CYBER SECURITY TRENDS

- Beyond Email: Multi-Channel Social Engineering
- The 2025 Ransomware Ecosystem
- From Recon to Narrative Control: Cyber's Operational Impact in 2025 Conflicts
- The Dominance of Chinese-Nexus Cyber Threats
- Unmonitored Devices: The Attackers' Launch Base

03 AI LANDSCAPE IN CYBER SECURITY

04 GLOBAL ANALYSIS

05 HIGH PROFILE VULNERABILITIES

06 2026 PREDICTIONS

07 RECOMMENDATIONS

08 AN EXPOSURE MANAGEMENT PERSPECTIVE

01

INTRODUCTION

INTRODUCTION

In 2025, the threat landscape evolved rapidly, becoming more interconnected and challenging to manage. Our analysis of global telemetry and incidents reveals a fundamental shift, marked by the emergence of new attack surfaces and techniques. Attackers are integrating AI, identity abuse, exposure exploitation, and ransomware into their campaigns.

The most significant change is the accelerated pace and scale at which attack opportunities are being executed. Data indicates that attackers are linking access, execution, and impact across various domains, from AI-driven social engineering and automation to the transformation of ransomware into a data-driven extortion economy. Edge devices and exposed infrastructure are increasingly used as launch points. These patterns were consistently observed across regions and industries in 2025, highlighting the swift combination of techniques to create tangible impacts.

AI exemplifies this transformation. This report views AI as a force multiplier that enhances targeting, scale, and adaptation in attacker activities, while also influencing risk prioritization and operational responses.

Our report is structured around attacker behavior and real-world data. The subsequent chapters delve into where attackers are focusing their efforts, how different techniques reinforce each other, and which exposure patterns most frequently lead to impact. This provides the necessary context to understand the trajectory of the threat landscape and what will be most critical in 2026.

I invite you to explore the data and findings presented in this report.

Lotem Finkelstein, VP Research

LOTEM FINKELSTEIN

VP Research

02

CYBER SECURITY TRENDS

BEYOND EMAIL: MULTI-CHANNEL SOCIAL ENGINEERING

Which attack surface is the easiest to exploit across all organizations? For attackers in 2025, the answer was the human component. In social engineering attacks, threat actors attempt to do just that: achieve compromise by manipulating human victims into providing the initial access for them. In such attacks, threat actors target employees, outsourced personnel, and third-party service providers to gain access to the organization's systems or sensitive information. Although these attacks are perceived to be less serious than those exploiting software or hardware vulnerabilities, they can be just as damaging as compromises achieved through other means.

For years, phishing emails served as the primary social engineering vector, and organizations became increasingly aware of these threats. However, by 2025, social engineering expanded beyond traditional email-based campaigns, adopting multi-platform, cross-channel, and highly targeted approaches that leverage phone calls, messaging applications, and real-time impersonation. At the same time, attackers have evolved how email and browser-based social engineering attacks are executed, shifting toward interaction-driven techniques such as ClickFix and its variants. These methods guide users through seemingly legitimate workflows designed to bypass security controls and inadvertently execute malware.

These approaches have resulted in millions of compromise attempts worldwide and contributed to several high-impact business breaches, resulting in significant financial losses for enterprises globally.

ClickFix: Social Engineering That Shifts Execution to the User

ClickFix emerged as one of the most significant social engineering techniques in 2025. First observed in 2024, ClickFix is an initial access method in which attackers manipulate users into executing malicious actions by presenting them with fraudulent instructions. These instructions, typically delivered through compromised or attacker-controlled websites, malvertising, or brand-impersonation emails, are crafted to resemble routine verification steps such as CAPTCHAs, validation checks, or error fixes. By appearing as legitimate steps required to continue normal activity, users are manipulated into running attacker-controlled content that ultimately delivers malware.

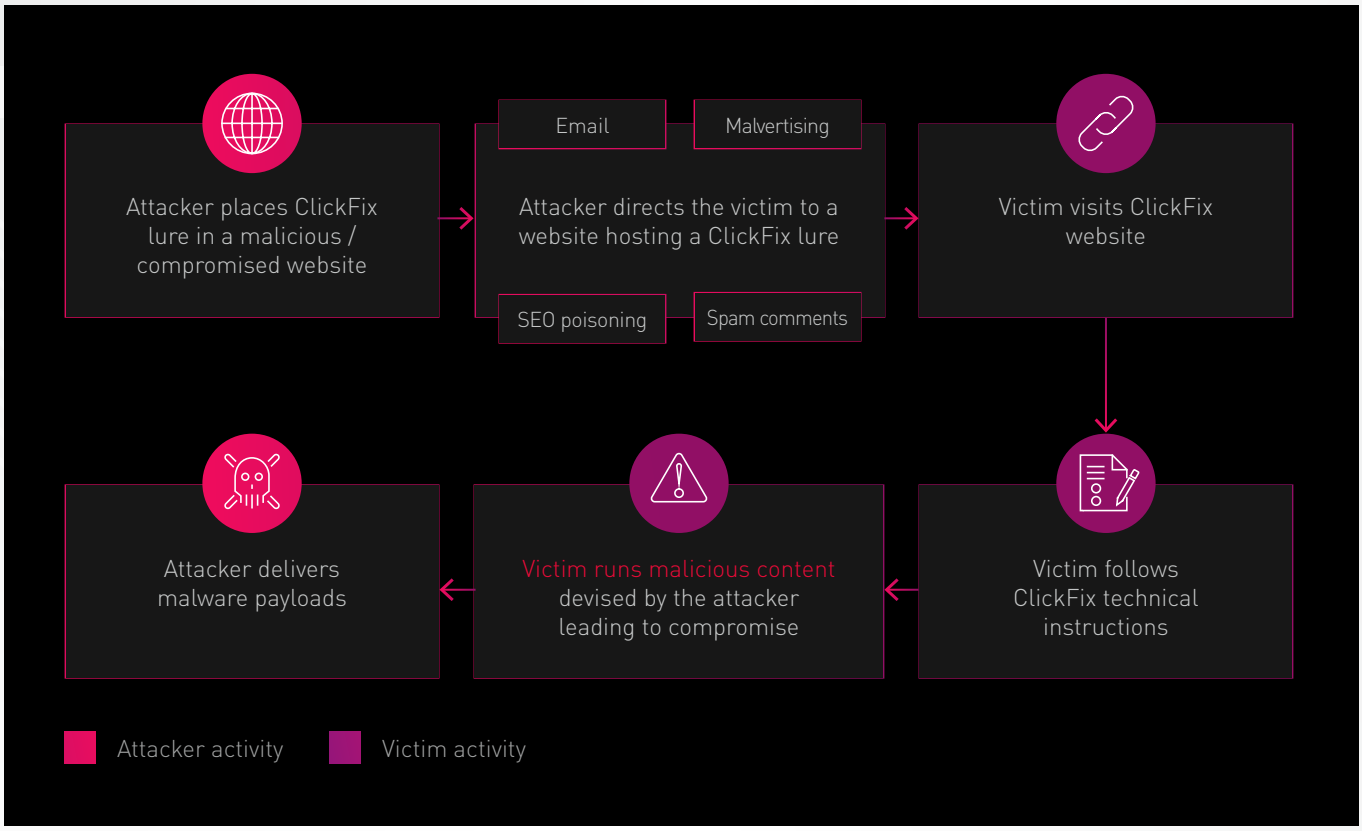


Figure 1: Flowchart of a ClickFix attack

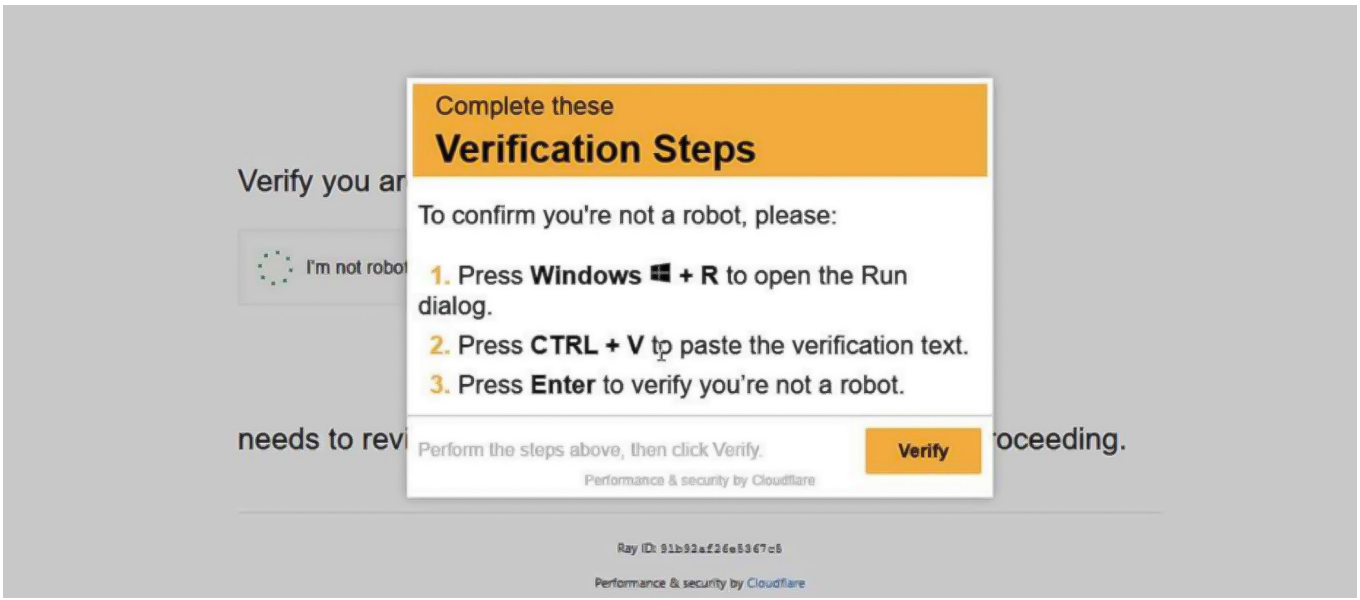
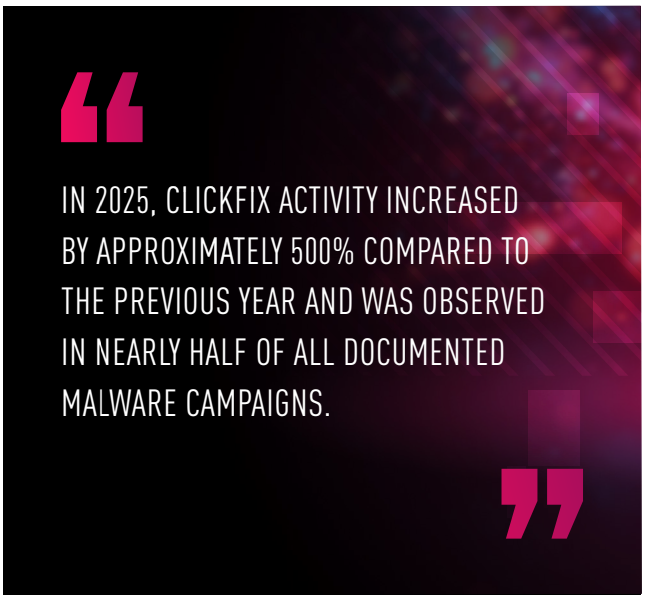


Figure 2: Example of a ClickFix website with a fake CAPTCHA prompt

This technique succeeds by exploiting user trust and the tendency to follow technical instructions. It has proven highly effective due to its simplicity, scalability, and ability to bypass certain security controls, as malicious actions are executed manually by the user rather than delivered through traditional file-based infection chains. As a result, its adoption has accelerated rapidly. In 2025, ClickFix activity increased by approximately 500% compared to the previous year and was observed in nearly half of all documented malware campaigns.



The technique has been widely adopted across the threat landscape, including by established cyber criminal groups behind major infostealer operations, such as RedLine and Lumma, and in attacks that delivered payloads leading to Interlock ransomware infections. At the same time, emerging malware families have increasingly used ClickFix as their initial attack vector. Recent examples include the MonsterV2 infostealer campaign targeting United States residents and a PureHVNC RAT campaign analyzed by Check Point Research. Beyond financially motivated crime, multiple nation-state-sponsored APT groups have also adopted ClickFix as a preferred delivery mechanism, in lieu of their more traditional initial access techniques.

ClickFix's success has led to the emergence of additional techniques that use the same social engineering approach. In mid-2025, threat actors began adopting FileFix, a ClickFix-derived technique that abuses legitimate operating system workflows to achieve initial access to the victim's device. FileFix relies on malicious or compromised websites to trigger a standard Windows Explorer window in which users are instructed to paste what appears to be a required

file path. This action causes attacker-controlled content to be executed, resulting in compromise without the use of traditional malware delivery methods. Originally introduced as a proof of concept, FileFix was weaponized by threat actors within weeks. Since then, multiple active campaigns have leveraged the technique to deliver malware payloads, including Interlock RAT and StealC infostealer, demonstrating how quickly successful social engineering methods transition from research to operational use.

In parallel, attackers have extended ClickFix's approach past code execution to account compromise. ConsentFix, which emerged toward the end of 2025, applies similar social engineering principles to cloud environments. It tricks users into completing a legitimate Microsoft/Azure OAuth login flow. It then instructs them to copy and paste a localhost URL that contains an OAuth authorization code into an attacker-controlled page. The stolen code is used to obtain tokens and gain access to the user's Microsoft account without capturing a password and completing multi-factor authentication (MFA).

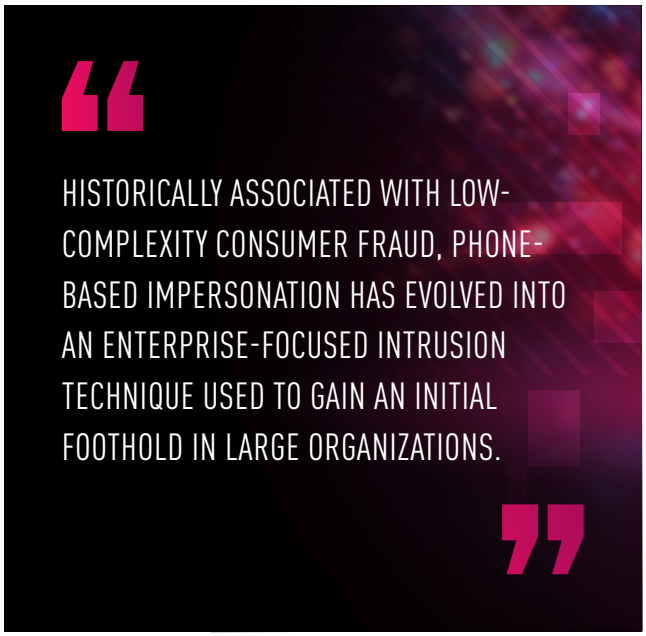
The popularity of ClickFix and its variants in 2025 has spread beyond Windows environments. Threat actors developed campaigns specifically targeting macOS users, as well as advanced threats that use ClickFix techniques to target Linux systems. As we saw with phishing kits, ClickFix began to commoditize by creating kits such as the IUAM ClickFix Generator, which enables attackers to create highly customizable, cross-platform ClickFix campaigns and rapidly adopt the technique at scale.

Coercing victims to initiate malicious activity on their own systems reflects a broader shift in attackers' social engineering strategies, which abuse user trust in legitimate processes across endpoints, browsers, and cloud identity platforms.

Voice-Based Social Engineering – The Weapon of Choice for Major Attacks

Voice phishing and impersonation gained significant traction in 2025, proving to be a highly effective means to exploit user trust. In these attacks, threat actors pose as trusted or authoritative figures and, following targeted reconnaissance, use rehearsed scripts to pressure victims to take actions such as resetting credentials, changing MFA codes, or granting network access. Historically associated with low-complexity consumer fraud, phone-based impersonation has evolved into an enterprise-focused intrusion technique used to gain an initial foothold in large organizations.

In 2025, voice-based impersonation became a preferred technique among highly sophisticated threat groups targeting major brands. These actors conducted in-depth reconnaissance, leveraged multiple communication platforms to engage victims, and executed complex, multi-stage social engineering scripts to achieve their goals. In several cases, voice-driven campaigns enabled attackers to gain initial access for some of the year's most damaging high-impact enterprise intrusions.



Most notably, this activity was associated with financially motivated threat actors such as Scattered Spider and the cluster commonly referred to as Scattered LAPSUS$ Hunters (SLH). Scattered Spider (also tracked as UNC3944 / Octo Tempest) is a highly effective, intrusion-focused cluster known for identity-centric initial access techniques, including help desk and IT vendor impersonation, MFA fatigue, and SIM-swap account takeover. SLH is a joint effort carried out by operators, tooling, and tactics associated with Scattered Spider, LAPSUS$, and ShinyHunters. These three groups have a track record of high-profile enterprise breaches and extortion. Notable past attacks include Shiny Hunters' breach of US telecom giant AT&T in 2024, for which the group received more than $350,000 in ransom payments; Scattered Spider's hack of MGM Resorts in 2023; and Lapsus$'s compromise of the identity authentication firm Okta via a breached third-party support provider back in 2022.

High-Impact Enterprise Incidents

SLH was linked to several high-impact incidents in 2025, where voice-driven social engineering served as the primary initial access vector against major enterprises, enabling data theft and extortion. In April 2025, Scattered Spider compromised British retailer Marks & Spencer's network through a targeted social engineering operation supported by extensive reconnaissance. The attackers gathered detailed information about the company's employees and internal processes, enabling them to convincingly impersonate a legitimate staff member when contacting a third-party help desk provider supporting Marks & Spencer. The attackers tricked a support engineer into resetting their password to gain access, leading to the deployment of the DragonForce ransomware. The incident forced Marks & Spencer to suspend online orders for over a month, disrupted in-store operations, and resulted in the theft of customer data. The company later estimated losses of approximately £300 million in lost profits and £136 million in direct incident response and recovery costs.

> IN 2025, VOICE-DRIVEN SOCIAL ENGINEERING EMERGED AS A PRIMARY INITIAL ACCESS VECTOR IN HIGH-IMPACT ENTERPRISE BREACHES, ENABLING DATA THEFT AND EXTORTION.

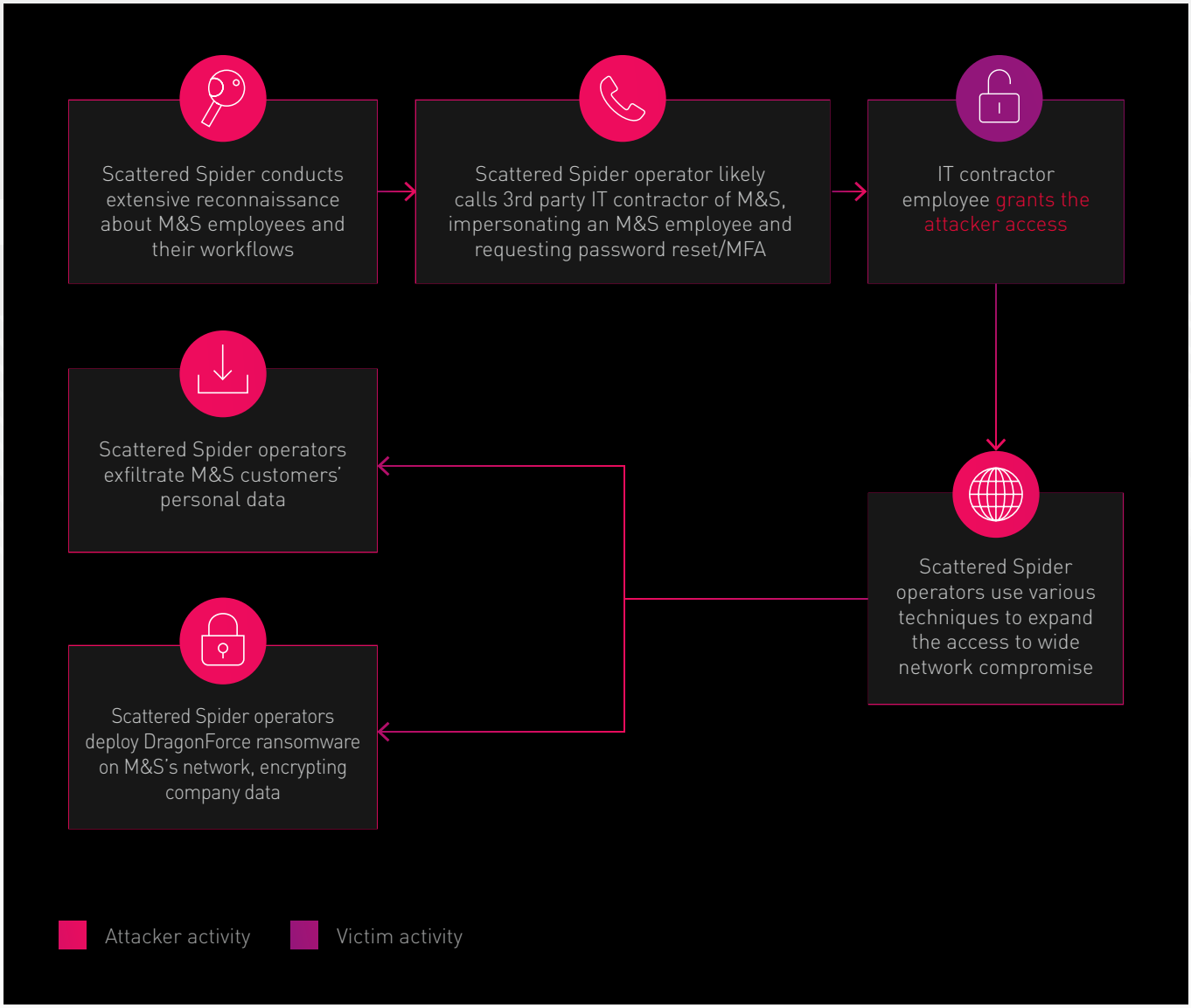


Figure 3: Scattered Spider's breach of Marks & Spencer

In another case, the British auto manufacturer Jaguar Land Rover (JLR) was targeted by SLH in August 2025. The attackers gained access to internal systems, exfiltrated customer data, and forced shutdowns across IT and manufacturing environments, resulting in disrupted production for several weeks. Although no technical assessment was published, reporting indicates that the intrusion likely involved social engineering techniques used against IT support teams, consistent with prior SLH activity. Estimated damages from the incident reached approximately £1.9 billion.

Earlier in 2025, ShinyHunters (also tracked as UNC6040), a threat actor known for sophisticated voice-phishing operations, conducted targeted campaigns against organizations' Salesforce environments to achieve large-scale data theft and extortion. The attackers focused on employees in English-speaking branches of multinational enterprises. They impersonated internal IT support staff to coerce victims into granting access or disclosing sensitive credentials, ultimately enabling data exfiltration from Salesforce instances. The threat actors later claimed the campaign affected approximately 40 organizations, including major global brands, and resulted in the exfiltration of nearly one billion records. These claims remain unverified.

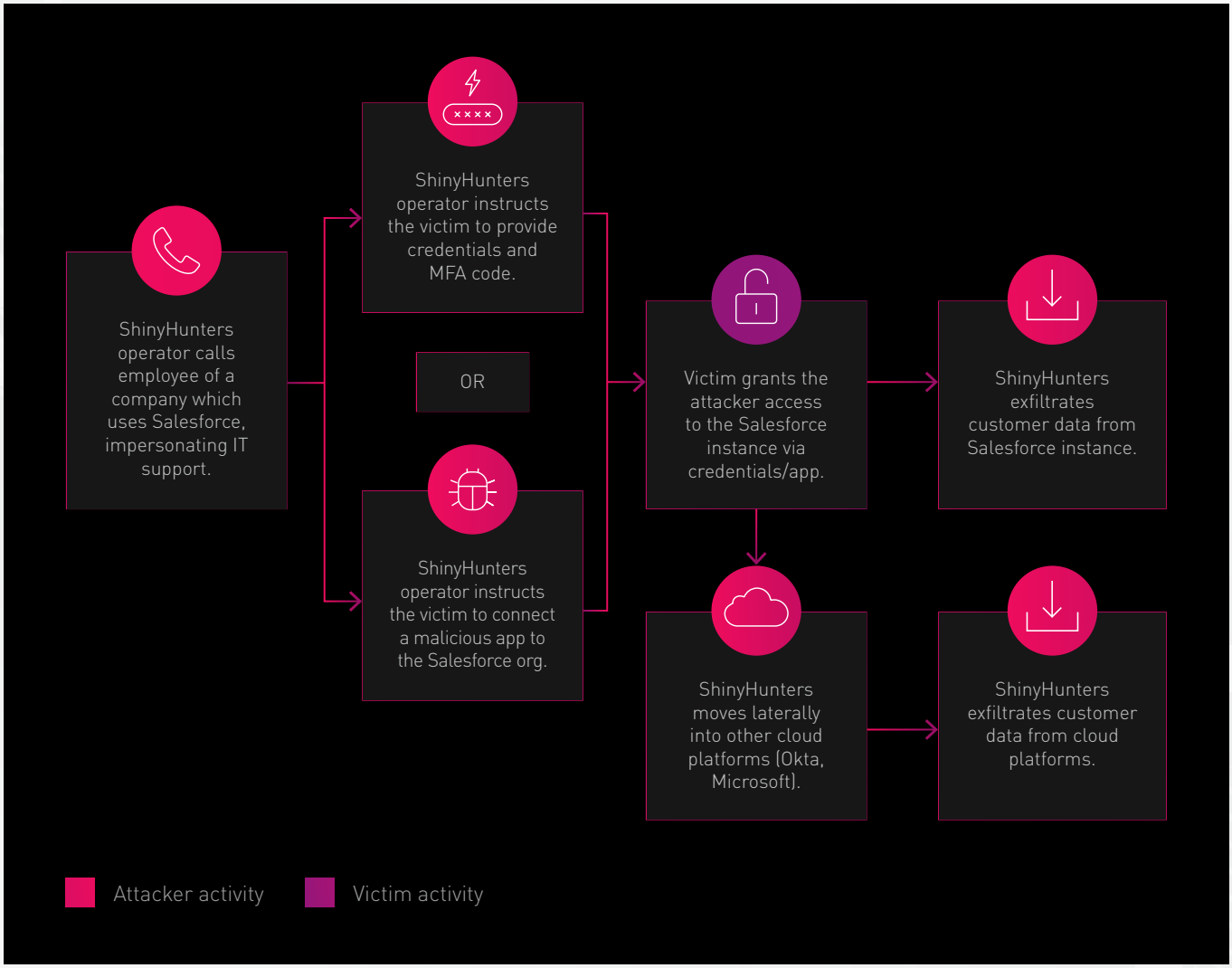


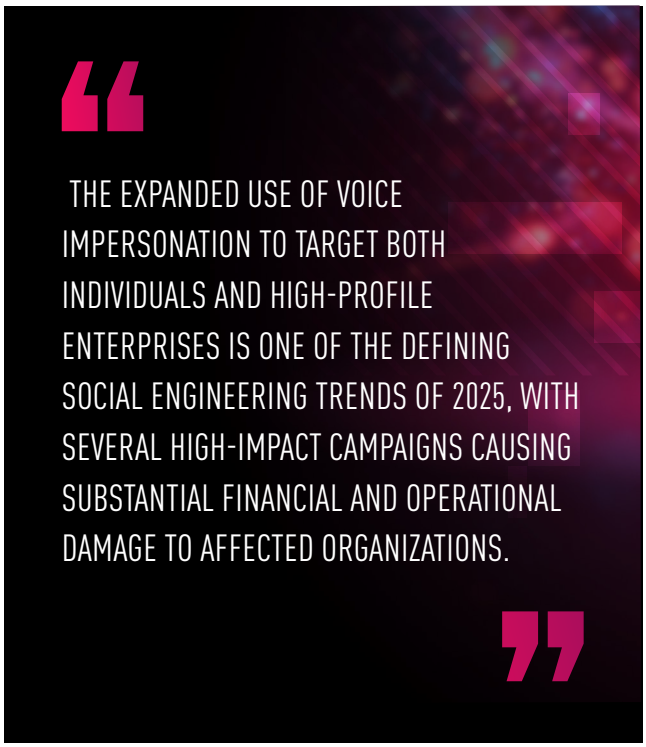
Figure 4: SalesForce attacks attributed to ShinyHunters

Available reporting indicates that in multiple cases, individuals affiliated with the SLH ecosystem are of Western origin. Public arrest records and indictments name individuals with United States, United Kingdom, and European citizenship. This may explain why these operators possess the linguistic fluency and cultural familiarity that enabled voice-based impersonation attacks against European and North American organizations.

Voice impersonation remains a significant fraud vector beyond enterprise intrusions, particularly in financially motivated scams targeting private citizens. In these cases, attackers commonly impersonate financial institutions or cryptocurrency platforms to pressure the victims into transferring funds or disclosing credentials that enable account takeovers. According to FBI reporting, the voice-enabled fraud and account takeover incidents in 2025 resulted in losses exceeding $250 million.

Voice Impersonation as a Defining Trend

The growing success of voice-based social engineering is driving increased demand for skilled impersonation operators within criminal ecosystems, as well as the emergence of a market for AI-driven voice impersonation tools and services. Overall, **the expanded use of voice impersonation to target both individuals and high-profile enterprises is one of the defining social engineering trends of 2025, with several high-impact campaigns causing substantial financial and operational damage to affected organizations.**



Victim-Initiated Social Engineering

In 2025, we observed a growing trend of victim-initiated (inbound) social engineering, where attackers deliberately steer targets into initiating contact, thereby increasing the perceived legitimacy of the interaction.

One campaign, identified as ZipLine by Check Point Research, involves attackers abusing organizations' public "Contact Us" pages to pose as legitimate business inquiries. This approach prompts employees, acting within their normal job responsibilities, to initiate follow-up correspondence with the attackers. The attackers then engage victims in weeks-long email exchanges before delivering a malicious ZIP attachment that deploys MixShell malware. The campaign has primarily been observed against manufacturing organizations.

HIGH-TRUST COMMUNICATION PLATFORMS ARE EMERGING SOCIAL ENGINEERING VECTORS

A similar victim-initiated pattern was observed in campaigns attributed to UNC6229, which targeted individuals in the marketing and digital advertising sectors to hijack corporate advertising and social media accounts. The actor created fake job postings on both legitimate platforms and attacker-controlled websites, relying on victims to initiate contact by applying for advertised roles. Initial communications were benign and personalized, building trust before shifting to malicious payload delivery or credential theft via phishing links.

While attacker-initiated phishing emails often have low success rates, reversing the interaction flow by engineering scenarios in which victims initiate or sustain communication significantly increases attacker's credibility and the likelihood of compromise.

Social Engineering Activity on Business Communication Platforms

Threat actors are increasingly expanding social engineering activity beyond email to messaging on social media platforms and messaging apps where user expectations and security controls are often weaker. These interactions typically mirror traditional phishing objectives such as coercing victims into executing malicious files or disclosing credentials, while benefiting from reduced user skepticism and a lack of dedicated security controls.

By operating through third-party messaging platforms, attackers can engage targets in more informal contexts, making it easier to build trust. This shift reflects a broader trend toward exploiting communication channels that fall outside traditional corporate security monitoring.

An example includes activity attributed to the Iranian APT group Nimbus Manticore, which was observed impersonating business professionals on LinkedIn to engage employees. Another Iranian APT group, Educated Manticore, leveraged messaging platforms such as WhatsApp for years as part of its social engineering methodology. This approach remains effective in 2025, as attackers build trust with their victims through informal communication channels while operating largely outside traditional enterprise security visibility and controls.

Lastly, threat actors are increasingly turning to enterprise collaboration platforms such as Microsoft Teams and Slack for social engineering channels. When organizational configurations allow external users to initiate chats or calls, these platforms provide attackers with a highly trusted environment in which to impersonate internal IT staff or service providers and engage employees directly via text, voice, or video.

Multiple campaigns observed over the past year leveraged Microsoft Teams as the initial access vector, with attackers messaging or calling employees from adversary-controlled Microsoft 365 tenants while posing as internal IT support. Victims are typically encouraged to install remote support tools, granting attackers full interactive access to their systems. This access is then abused to deploy next-stage malware, such as the Matanbuchus loader, and eventually achieve network-wide compromise, which may also involve ransomware. **These campaigns highlight how collaboration platforms, designed to streamline business communication, are**

increasingly being abused as high-trust attack surfaces, enabling attackers to bypass standard email defenses and exploit human weaknesses to achieve rapid compromise.

Navigating the Rapid Evolution of Social Engineering Threats

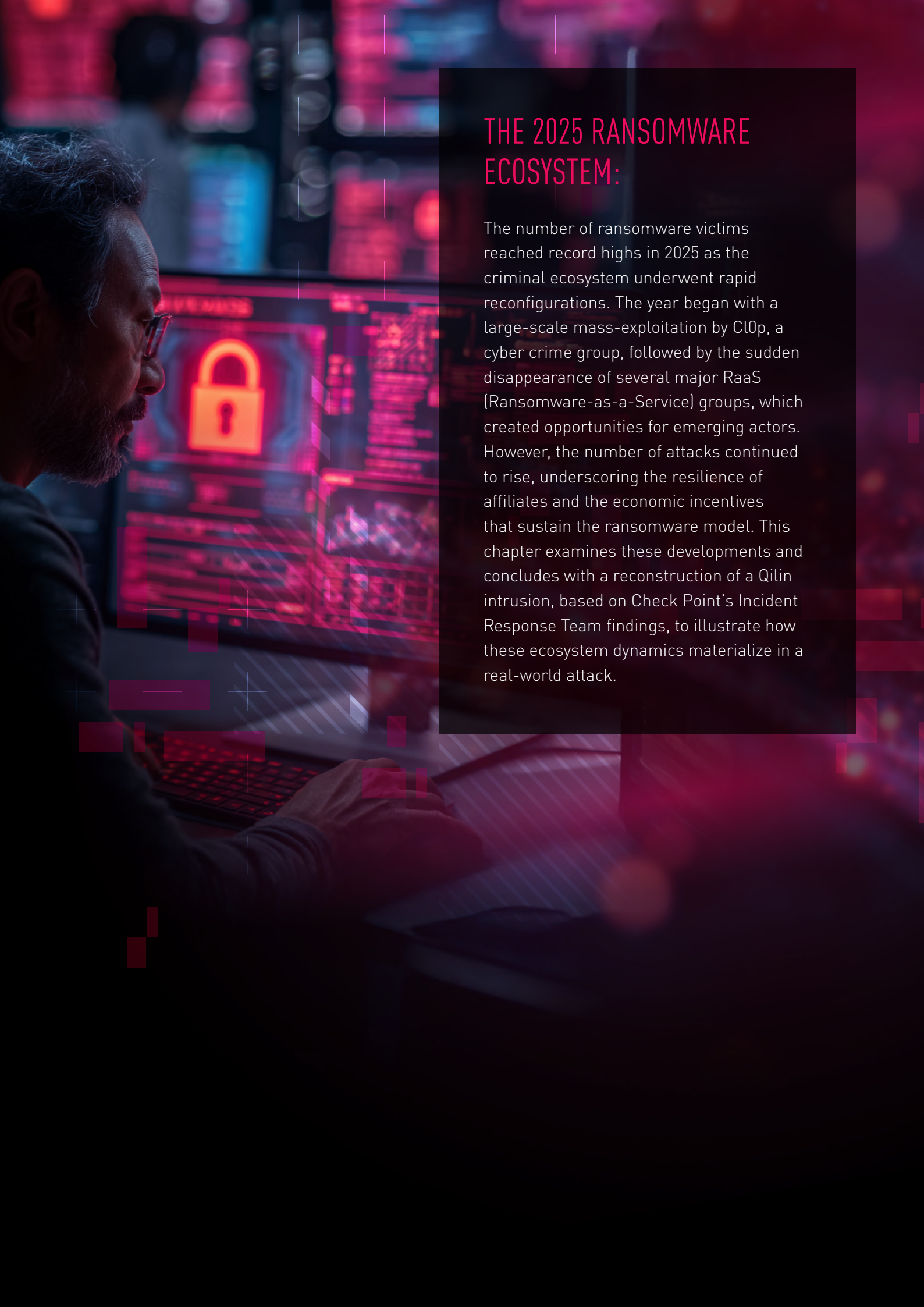
In 2025, social engineering took center stage as the dominant attack vector across the threat landscape, from scams and opportunistic malware campaigns to the most damaging enterprise compromises. Threat actors expanded their techniques, increasingly leveraging multiple platforms, diverse psychological tactics, and creative technical approaches. Tactics such as ClickFix and voice impersonation proved especially effective, becoming the primary tools for leading malware and intrusion groups.

As noted earlier, the human element remains the weakest link in organizational security. In 2026, social engineering activity is expected to intensify further. Generative AI is lowering the barrier to highly convincing attacks, while the rapid adoption of new tools and solutions provides threat actors with an expanding set of trusted workflows to exploit. As a result, social engineering represents a growing, adaptive threat that organizations must treat as a central security challenge.

> Social engineering expanded beyond traditional email-based campaigns, enabling multi-platform, cross-channel, and highly targeted approaches that leverage phone calls, messaging applications, and real-time impersonation.

SERGEY SHYKEVICH

Group Manager
Threat Intelligence



THE 2025 RANSOMWARE ECOSYSTEM:

The number of ransomware victims reached record highs in 2025 as the criminal ecosystem underwent rapid reconfigurations. The year began with a large-scale mass-exploitation by Cl0p, a cyber crime group, followed by the sudden disappearance of several major RaaS (Ransomware-as-a-Service) groups, which created opportunities for emerging actors. However, the number of attacks continued to rise, underscoring the resilience of affiliates and the economic incentives that sustain the ransomware model. This chapter examines these developments and concludes with a reconstruction of a Qilin intrusion, based on Check Point's Incident Response Team findings, to illustrate how these ecosystem dynamics materialize in a real-world attack.

2025 was defined by rapid turnover among the top ransomware groups, the rise of actors such as Qilin, and the re-emergence of established brands like Cl0p and LockBit, all against a backdrop of growing global policy debates over ransom payments, reporting mandates, and the limitations of law enforcement disruption. Ransomware operations increasingly incorporate AI into different stages of the attack lifecycle, including malware development, stolen-data analysis, legal and regulatory assessment, and support for negotiation and extortion activities.

Ransomware activity reached unprecedented levels in 2025. Over 7,960 victims were named on data-leak sites operated by double-extortion groups, a 53 percent year-over-year increase. Q1 recorded 2,289 published victims, a 134 percent YoY increase, driven in part by Cl0p's exploitation of zero-day vulnerabilities. This made Q1 the most active quarter ever recorded in our dataset, a record that was subsequently surpassed in Q4 with 2,473 published victims.

Figure 1 illustrates the sustained, multi-year upward trajectory in ransomware victims.

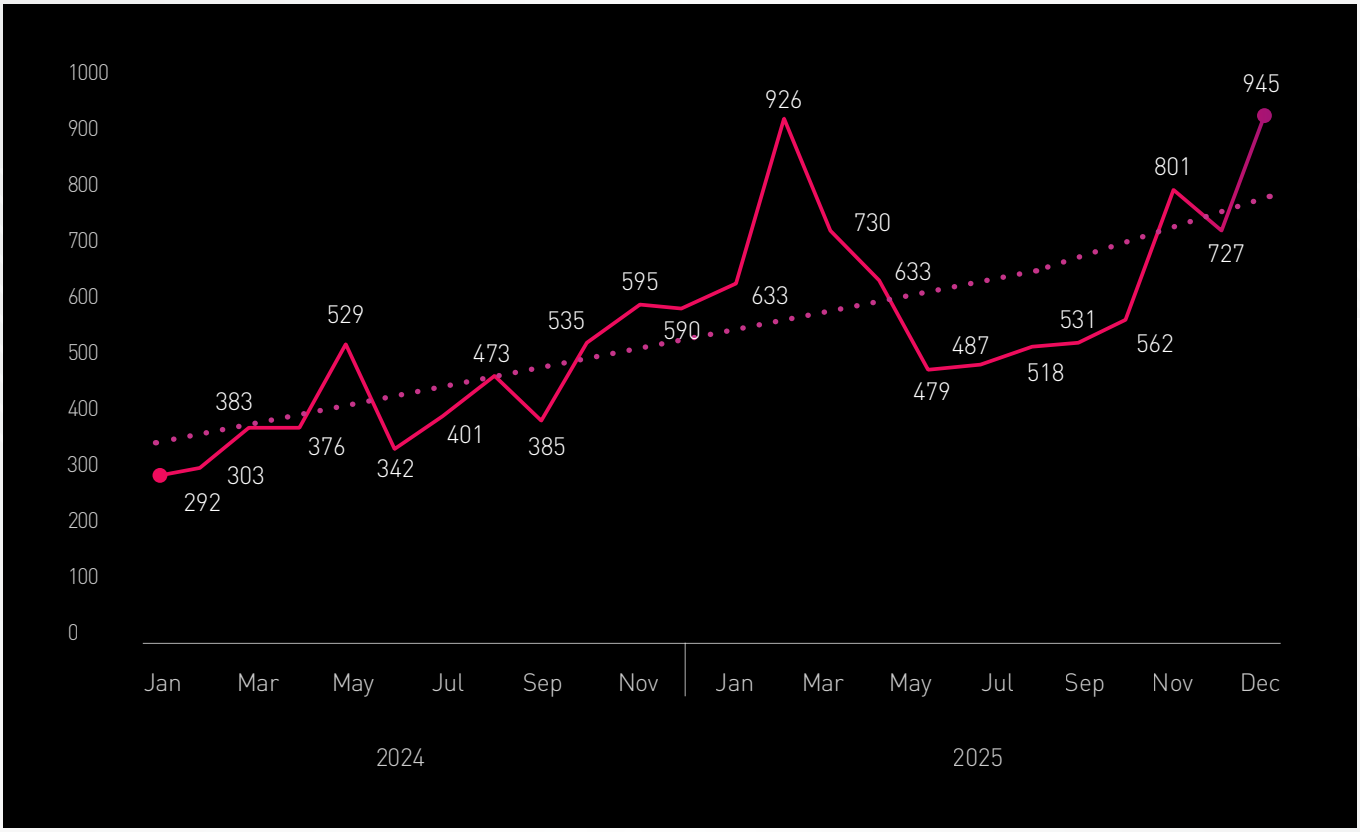


Figure 1 - Published ransomware victims per month

Mid-Year RaaS Disruption and Affiliate Realignment

In Q2, several high-profile RaaS programs abruptly disappeared, while victim count remained well above 2024 baselines. 8Base and Phobos were disrupted by coordinated international law enforcement operations that seized leak and negotiation sites, resulting in multiple arrests of key operators and affiliates. Other prominent ransomware groups, including BianLian, Hunters, and Cactus, either rebranded and shifted entirely to data extortion models or quietly ceased publishing new victims. RansomHub, which posted more than 760 victims since it appeared in 2024, went offline without warning in early April 2025. This wave of exits temporarily left many affiliates, the operators who conduct the attacks, without a stable RaaS brand. Qilin and DragonForce moved quickly to fill the gap, actively recruiting former RansomHub and LockBit affiliates on criminal forums. By Q3, both ranked among the most active data-leak site operators.

Qilin emerged as the primary beneficiary of the mid-year shake-up. Its volume of published victims increased steadily through Q2 and Q3, driven by its success in attracting unaffiliated intrusion operators following the collapse of several long-standing groups. By mid-2025, Qilin was among the most active RaaS operators, surpassing many legacy brands.

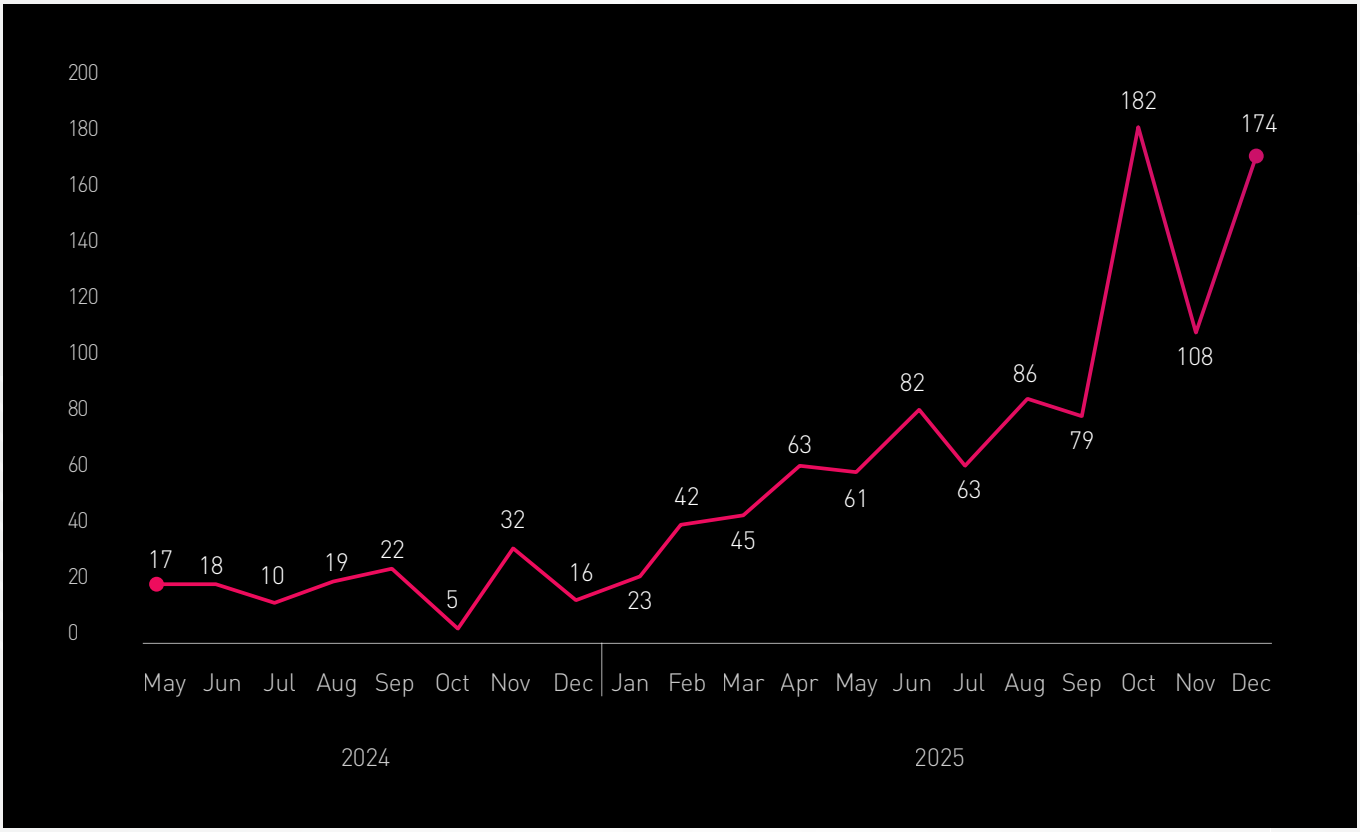


Figure 2 - Monthly count of Qilin's published victims

Proliferation of Independent Double-Extortion Groups

2025 also saw an unprecedented proliferation of small, independent, double-extortion groups. At the end of 2024, approximately 90 identifiable brands were publishing victims on data-leak sites (DLS), while 2025 recorded 140 distinct groups, an increase of more than 50 percent. This illustrates how smaller operations quickly filled the vacuum created by the retreat of major RaaS programs. Many of the new actors operated without formal affiliate programs, instead relying on single teams or small partnerships that could launch attacks without the overhead, revenue sharing, or infrastructure demands of larger RaaS frameworks.

However, toward the end of 2025, the landscape shifted again. Larger, better-known brands like Qilin reasserted their earlier dominance. Akira's activities surged, and Cl0p reemerged after months of near silence, continuing its pattern of high-impact, opportunistic mass-exploitation campaigns. The reappearance of LockBit, now branded as LockBit 5.0, further signaled a return of the major RaaS groups.

This cycle, in which dominant groups disappear and smaller actors proliferate, only to later coalesce around a few large players, illustrates the structural role RaaS programs play in the ransomware ecosystem. Successful extortion depends on the victim's belief that the threat actor will both decrypt data and refrain from leaking it once the ransom is paid. An affiliate association with a recognizable RaaS brand lowers transaction friction and increases the likelihood of payment. However, increased visibility exposes RaaS operators to increased pressure from law enforcement. Operators can reduce the attention they attract through cyclical rebranding, shutting down one name and infrastructure set, and reappearing under a different one. This dynamic complicates attribution and disruption efforts for defenders and investigators, and helps explain the recurring rise, fall, and reappearance of dominant RaaS brands observed throughout 2025.

“THIS CYCLE, IN WHICH DOMINANT GROUPS DISAPPEAR AND SMALLER ACTORS PROLIFERATE, ONLY TO LATER COALESCE AROUND A FEW LARGE PLAYERS, ILLUSTRATES THE STRUCTURAL ROLE RAAS PROGRAMS PLAY IN THE RANSOMWARE ECOSYSTEM”

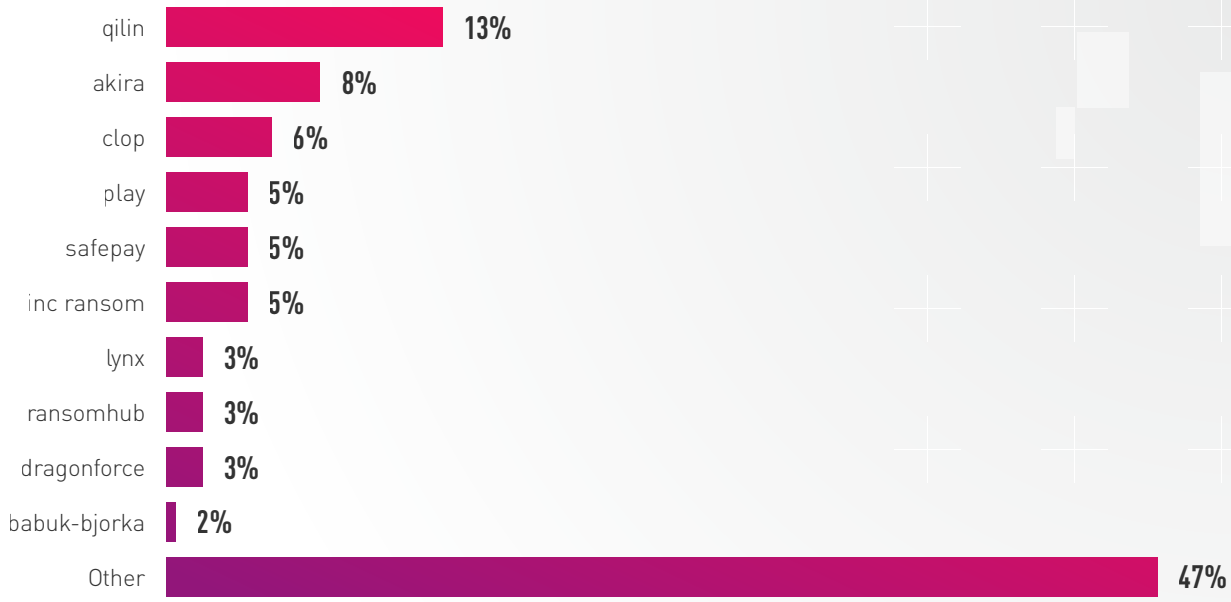


Figure 3 – Top 10 RaaS groups in 2025 by percentage of published victims

Large RaaS Operators Continue to Drive Volume

A review of the most active ransomware groups in 2025 shows that the ecosystem remains anchored by long running RaaS operations, despite the influx of new brands. Several of the year's top actors, such as Qilin and Play, active since 2022; Akira, Inc Ransom, and DragonForce, which emerged in 2023; and Lynx and RansomHub, established in 2024, maintained multi-year continuity, strong affiliate networks, and stable infrastructure. Their sustained presence underscores that while smaller independent groups proliferated throughout 2025, large RaaS programs continued to drive the overall attack volume.

Qilin – The Emerging Dominant RaaS Group

Qilin emerged as the dominant RaaS group of 2025, publishing the identities of over 1,000 victims on its DLS after they refused to pay a ransom. The group, active since 2022, was well positioned to capitalize on the temporary vacuum caused by the disappearance of major rivals, such as RansomHub, and the inactivity of LockBit. Following RansomHub's abrupt collapse in April 2025, Qilin aggressively recruited orphaned affiliates. As a result, the group's monthly victim counts nearly tripled during 2025, increasing from an average of roughly 35 victims per month in Q1 to over 150 in Q4. Qilin was the single most active ransomware group in 2025, according to DLS data, consistently ranked first or second in monthly victim disclosures, with activity levels exceeding those of Akira, DragonForce, and Play.

RaaS Platform Capabilities and Extortion Model

Qilin operates a fully featured RaaS framework that provides affiliates with an administrative panel supporting the end-to-end attack lifecycle. The platform includes access to the encryptor, leak infrastructure, payment negotiation tooling, and operational support. The group's extortion methodology reflects broader ecosystem trends toward data-exfiltration operations: negotiations rely more heavily on the threat of regulatory exposure, reputational harm, and operational disruption than on decryption alone.

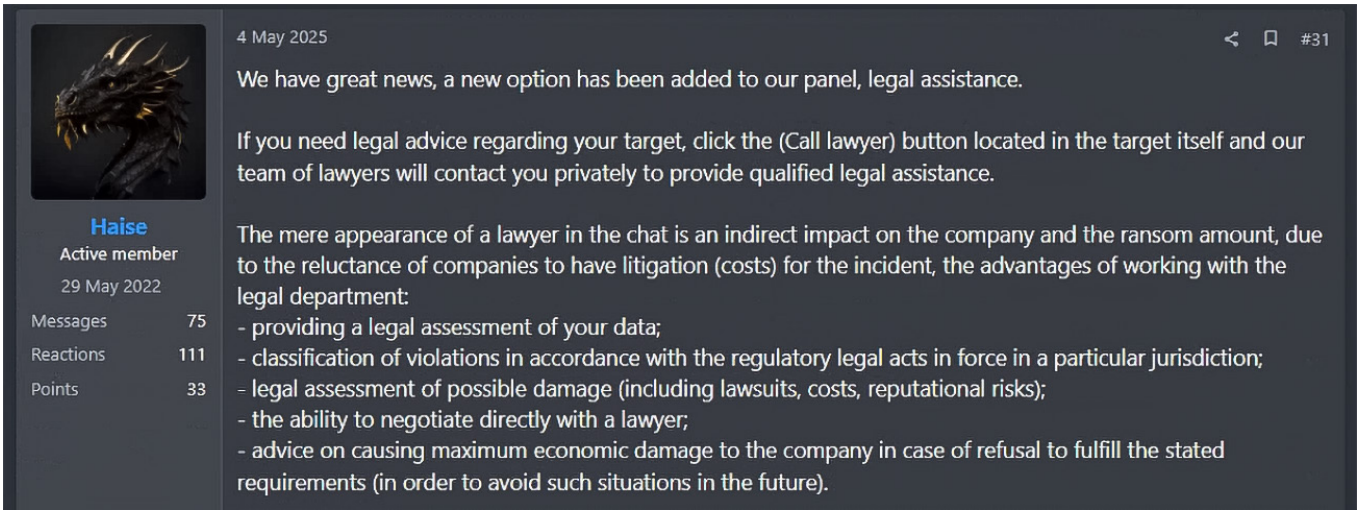


Figure 4 – Qilin's promotion of new extortion tools in a Dark Web forum

To increase the leverage affiliates can exert on victims, Qilin introduced a set of enhanced "pressure" services. These include a so-called legal review capability, in which stolen data is examined for indicators of compliance or mandatory disclosure requirement violations, with documentation intended for submission to agencies such as tax authorities or the FBI. The group also advertises tools for mass messaging a victim's corporate email accounts and phones, as well as public-facing leak narratives through purported "journalistic" intermediaries. Several of these functions resemble GenAI-supported content generation, which may indicate the integration of automated drafting and distribution tools for extortion.

Targeting Profile and Affiliate Incentives

Despite presenting themselves as "idealists" driven by patriotic motives, Qilin's targeting is global (excluding CIS countries), sector-agnostic, and clearly financially motivated. The group offers affiliates a competitive profit share of 80 to 85 percent, positioning itself as a high-margin alternative to other RaaS groups and reinforcing its attractiveness during a year marked by significant affiliate displacement.

Cl0p – A Zero-Day Outlier

Cl0p shaped both the beginning and end of the 2025 ransomware timeline. Unlike traditional RaaS actors, Cl0p consistently relied on highly strategic zero-day exploits of widely used enterprise software to simultaneously compromise hundreds of organizations. Cl0p extortion operations are based solely on the threat of publishing stolen data, rather than file encryption.

Their first major campaign of the year targeted Cleo's LexiCom, VLTrader, and Harmony file-transfer applications via two unauthenticated remote-code-execution vulnerabilities. The February campaign resulted in over 335 publicly reported victims, primarily across North American manufacturing, retail, logistics, and supply-chain operators. Much of the record-high victim count in Q1 was driven by this event.

Cl0p's second major operation surfaced in Q3-Q4 when multiple Oracle E-Business Suite zero-day vulnerabilities were exploited in the wild, including CVE-2025-61882 and CVE-2025-61884. Investigations revealed that exploitation began as early as August, months before vendors released patches. The high-profile victims included

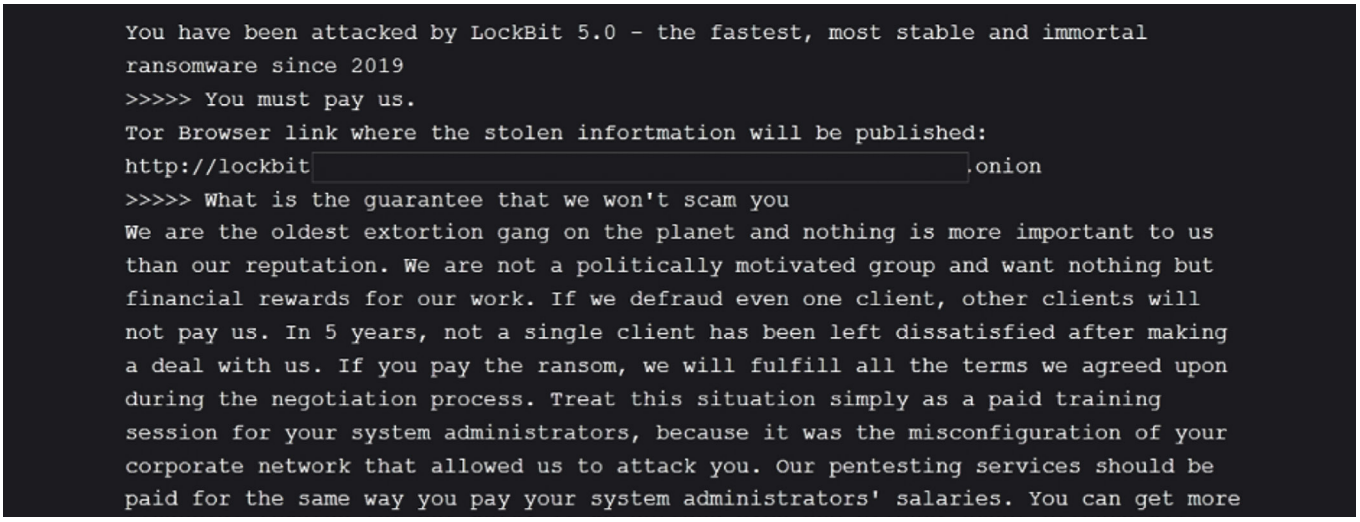


Figure 5 – A ransom note from a LockBit 5.0 attack in mid-September 2025

universities, airline subsidiaries, major media organizations, and multinational manufacturers. The campaign triggered additional activity when the exploit code was publicly leaked in October, enabling other threat actors to replicate the attacks.

The Return of LockBit

After suffering near-complete operational collapse following the Cronos law enforcement operation in early 2024, LockBit spent the first half of 2025 largely inactive. Leak site publications fell to fewer than five monthly victims. However, the group's administrator, LockBitSupp, repeatedly signaled an imminent return in the underground forums.

In September 2025, LockBit officially relaunched as LockBit 5.0, with an updated encryptor, enhanced evasion capabilities, and a redesigned affiliate interface. The group immediately resumed active intrusions, primarily targeting US organizations. In December, victim publication resumed, with over 100 victims in its first month of renewed activity.

LockBit's return demonstrates that many affiliates still prefer to work under a recognized and stable group when one is available. RaaS groups that maintain operational security and preserve affiliate trust may attract enough participation to shift the ecosystem back toward a model dominated by a small number of major groups.

Constraining Incentives: Payment Restrictions and Compulsory Reporting

The sustained growth of ransomware victims in 2025 demonstrates that law enforcement takedowns of major RaaS groups, while disruptive, failed in reducing overall attack volume. Affiliates routinely cluster under new group names or migrate to alternative platforms, leading governments to shift focus toward constraining the financial incentives that sustain the ransomware ecosystem.

In 2025, the United Kingdom advanced comprehensive proposals, including a potential ban on ransom payments by public sector institutions and mandatory reporting. The European Union's NIS2 Directive introduced strict incident reporting timelines requiring

organizations to disclose ransomware incidents and, in many cases, payment status. Australia's Cyber Security Act 2024, effective in June 2025, established the world's first national mandatory ransomware-payment reporting framework, requiring detailed disclosures within 72 hours of an extortion attempt. While no federal nationwide ban exists in the US, OFAC (Office of Foreign Assets Control) sanctions restrict payments to designated ransomware groups, and several states prohibit or mandate disclosure of public-sector ransom payments. Collectively, these measures signal a policy shift toward reducing ransomware profitability through transparency and payment restrictions, rather than relying solely on law enforcement activity.

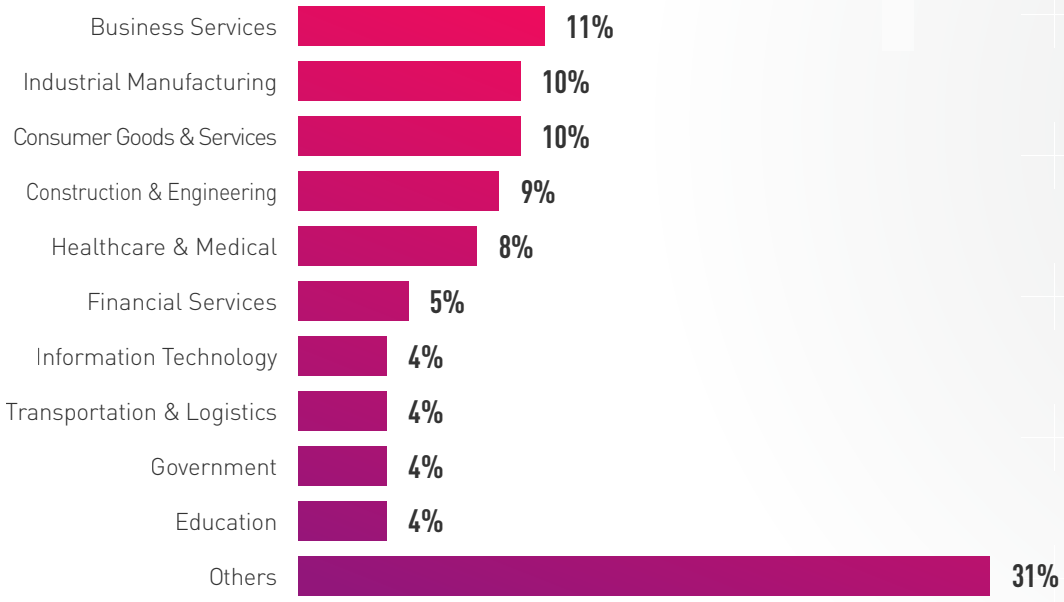


Figure 6 – The percentage of ransomware victims by industry

According to DLS data, commercial sectors such as business services, consumer goods & services, and industrial manufacturing remain the most frequently targeted, while government and education rank far lower among victims. This profile contrasts sharply with the sector distribution typically observed in broader cyber attacks and likely reflects differences in ransom payment behavior: public sector and educational institutions are generally less willing or legally unable to pay ransoms, reducing their attractiveness to financially motivated actors. This pattern aligns with the regulatory trends discussed above in which governments are increasingly restricting or discouraging ransom payments by public entities.

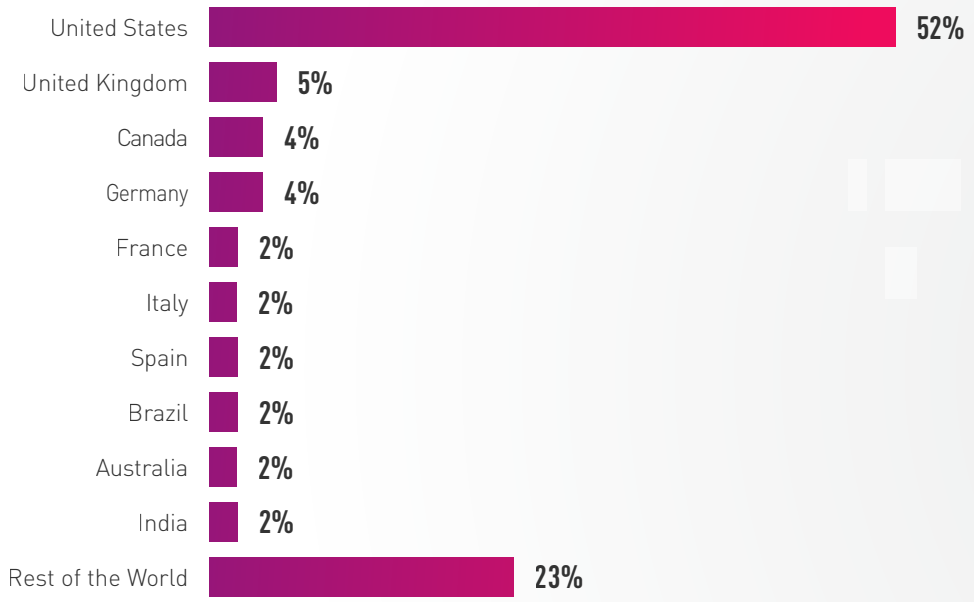


Figure 7 – The percentage of ransomware victims by country

Geographically, the distribution of victims in 2025 remains heavily concentrated in the US, which accounts for approximately 52 percent of all disclosed cases. The United Kingdom follows at 5 percent, with Canada and Germany each at 4 percent. Within the RaaS model, affiliates generally choose their own targets, resulting in a geographic distribution that reflects broader economic and exposure patterns rather than the strategic preferences of any specific RaaS. While operators may impose specific prohibitions, such as avoiding organizations in former Soviet republics, non-profit entities, or healthcare providers, these restrictions shape only a minority of campaigns.

THE CHECK POINT INCIDENT RESPONSE TEAM: INSIDE A QILIN RANSOMWARE ATTACK

Qilin was the most prolific RaaS group of 2025. The following case study examines a significant attack against a Western European electric power company, illustrating how a single misconfiguration, compounded by weak identity protection, monitoring, and access privileges, allowed a Qilin affiliate to gain complete control over an enterprise environment and execute a destructive, double-extortion attack. We reconstructed the event from endpoint telemetry, VPN and RDP logs, and artifacts left behind by the attackers. We highlight not only their conduct, but also the operational realities faced by the incident response team and CISO after encryption.

Figure 8 - Timeline of the Qilin attack (DT indicates the encryptor deployment date)

Initial Compromise: A Privileged Account on BYOD (Bring Your Own Device)

Ransomware attacks typically contain multiple phases, culminating in the deployment of the encryptor, and in this instance, Qilin is used. The incident began with a "super" domain administrator account referred to here as ADMIN. The account held extensive privileges and was used routinely for daily management tasks. Critically, it could be accessed from an unmonitored personal laptop over VPN and did not require multi-factor authentication (MFA).

FAILURE POINT 1:
VPN from unmanaged BYOD without MFA

In early September 2025, the attacker gained immediate high-privilege access through a seemingly legitimate VPN login using valid ADMIN credentials. No brute force or vulnerability exploitation was required. During the investigation, multiple indicators suggested the credentials were harvested from the laptop; however, because that was outside corporate control, it could not be examined.

FAILURE POINT 2:
ADMIN account operated from unmanaged device

Minutes after the VPN connection, the attackers wrote 1.exe, a copy of Mimikatz, into the C:\PerfLogs directory of the primary domain controller. The directory was neither monitored nor restricted, but Sysmon logs were already deployed and captured the creation and execution of the file. With domain admin privileges already in hand, the attackers did not require further escalation and began immediate reconnaissance.

Mapping the Environment: Quiet Reconnaissance and Host Discovery

FAILURE POINT 3:
Unmonitored & unrestricted directory allows malicious tools installation

Over the next hour, attackers executed PowerShell-based DNS queries, enumerating thousands of hostnames to identify active systems and rapidly build a near-real-time map of file servers, hypervisors, backup appliances, and management nodes. A brief query of the "net local" group administrators confirmed that the compromised account had unrestricted access.

While operational staff detected no anomalies, the attackers had already assembled a clear operational view of the environment.

Controlled Lateral Movement and the Backup Problem

During the next few days, the attackers moved laterally with precision and restraint, accessing a central file server, the primary backup server, and an IT management/jump server via RDP. Only the legitimate ADMIN account was used, with native tools and privileged credentials.

The targets were selected deliberately. Backup infrastructure and management servers were singled out early on to ensure that, in the event of encryption, recovery would be slow, painful, and uncertain.

Dormant Dwell Time: Five Weeks of Invisible Exposure

For more than a month, the attackers' activity was kept to a minimum despite retaining full access. No continuous scanning, persistence mechanisms, or repeated logins were observed.

From an incident response perspective, this long period of silence is particularly dangerous as attackers use this time to complete discovery, map the environment, and position themselves to deliver maximum operational impact. By the time visible disruption occurs, the attack has been underway for a while.

Data Exfiltration: MEGAsync on the File Server

On October 8, one day before encryption, attackers installed and ran MEGAsync on the central file server for several hours, after which it was removed. Network telemetry revealed correlated outbound data transfers, though insufficient to enumerate the content. The pattern aligns with a common double-extortion workflows: staging the data, exfiltrating it to cloud storage, removing the tool, and then preparing for encryption.

FAILURE POINT 4:
Unrestricted download and installation of file-sharing app

While the incident response team could not conclusively prove what was stolen, the timeline and tooling strongly indicate data exfiltration, a critical factor for regulatory reporting and ransom negotiation strategy.

The Night Before Impact: Backup Destruction and Payload Deployment

On October 9, the attackers reconnected through the jump server. They accessed a backup appliance using the ADMIN account, executing destructive commands against file systems, volumes, and services that degraded backup integrity.

FAILURE POINT 5:
Unmonitored after-hours destructive access to bkp server.

FAILURE POINT 6:
No cold backups

This activity was anomalous, involving after-hours access to backup infrastructure and destructive actions by a privileged account, yet no alerts were generated or escalated. Controls existed, but monitoring was absent.

Execution: Qilin Deploys via PerfLogs

Near midnight on October 9, the attackers re-executed C:\PerfLogs\1.exe, this time deploying the Qilin ransomware payload rather than Mimikatz. Reusing the filename helped the operator blend in with earlier artifacts, delaying detection as encryption began.

```
=== END COMMANDLINE CONFIGURATION ===
[INFO] Current execution context: [redacted]
[INFO] Set SeDebugPrivilege successfully.
[INFO] Set SeImpersonatePrivilege successfully.
[INFO] Set SeIncreaseBasePriorityPrivilege successfully.
[DEBUG|MUTEX] Trying to lock mutex
[INFO|MUTEX] Ownership of mutex taken successfully
[INFO] Gone into background. You can close this console window now.
[INFO] Successful change of ErrorMode
[DEBUG] Current exe path: "C:\\PerfLogs\\1.exe"
```

Figure 9 – A Qilin log showing the execution user and path

```
=== END COMMANDLINE CONFIGURATION ===
[INFO] Current execution context: [redacted]
[INFO] Set SeDebugPrivilege successfully.
[INFO] Set SeImpersonatePrivilege successfully.
[INFO] Set SeIncreaseBasePriorityPrivilege successfully.
[DEBUG|MUTEX] Trying to lock mutex
[INFO|MUTEX] Ownership of mutex taken successfully
[INFO] Gone into background. You can close this console window now.
[INFO] Successful change of ErrorMode
[DEBUG] Current exe path: "C:\\PerfLogs\\1.exe"
```

Figure 10 – A Qilin log showing the fingerprinting of the machine

Forensic artifacts provided exceptional visibility into its internal workflow, including:

- **Environment fingerprinting** to detect analysis environments by checking CPU and platform characteristics (Figure 10)
- **Domain and share enumeration** through Active Directory queries to list all domain-joined systems
- **Multithreaded remote encryption** over SMB and admin shares, avoiding binary deployment to each endpoint
- **Anti-forensics routines,** including wiping logs and deleting binaries
- **Persistence creation** via autorun registry keys to relaunch 1.exe after reboot

FAILURE POINT 7:
No detection of mass-file encryption

On the central file server, a mid-run reboot interrupted local encryption. On the backup server, the ransomware ran uninterrupted for hours, encrypting broad swaths of infrastructure, including VMs, branch servers, and key application systems.

By the morning of October 10, as staff arrived to find systems locked and ransom notes, the technical phase of the attack had already concluded.

The Morning After: Incident Response Triages Under Uncertainty

When the incident response team assembled, the organization faced a convergence of worst-case

conditions: core virtual infrastructure was encrypted, backups were partially destroyed or unreliable, branch operations were disrupted, ransom notes appeared on multiple endpoints, and the SIEM (Security Information and Event Management) was also encrypted, eliminating a central log source.

The CISO faced immediate questions: Was sensitive data exfiltrated? Were attackers still inside the environment? Were any backups intact to support restoration? Should the incident be treated as a recovery operation, a data breach, or both?

Surviving Sysmon logs from several hosts enabled investigators to reconstruct the intrusion and determine that attacker access likely ended.

FAILURE POINT 8:
No proactive IR program and playbook

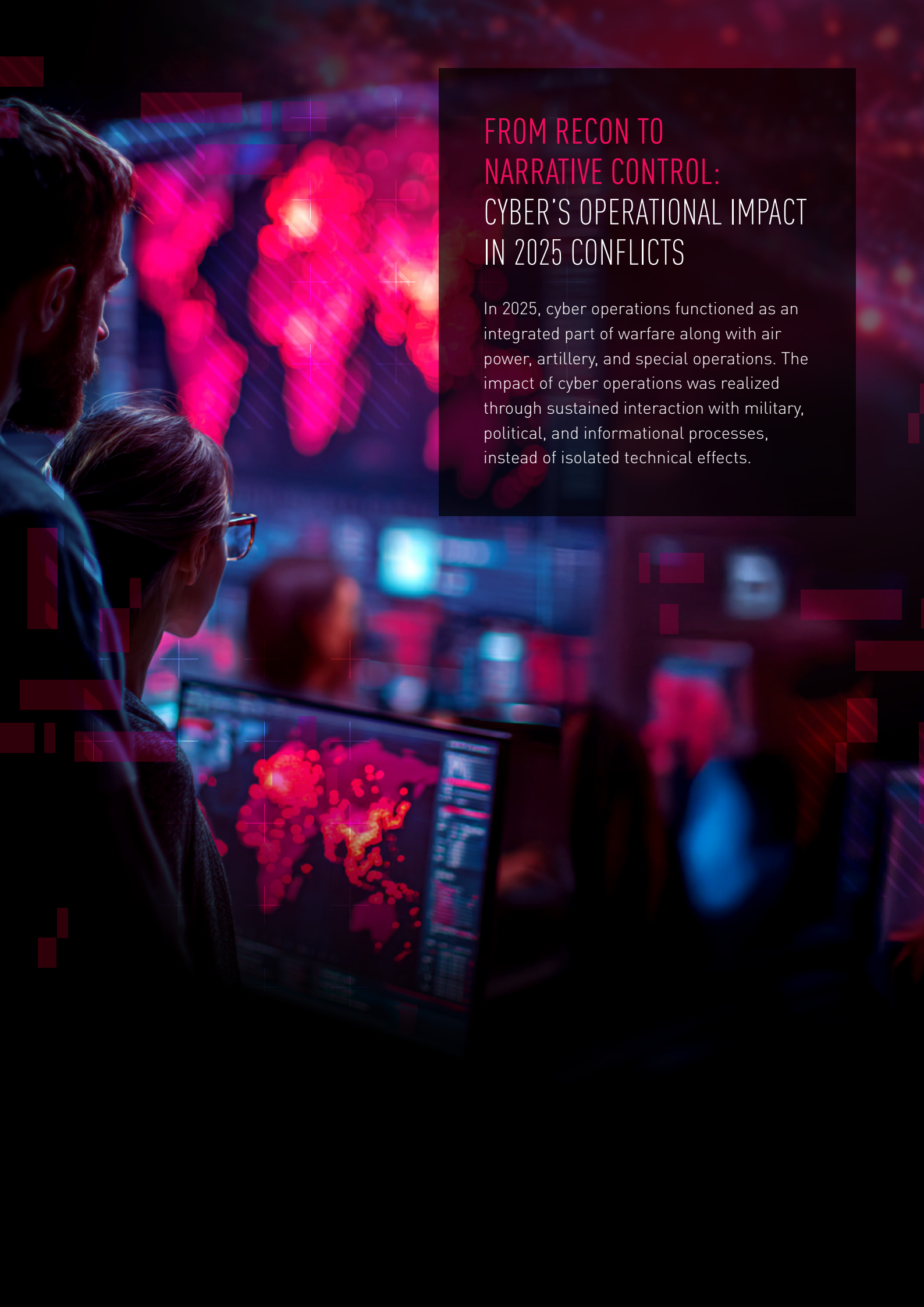
Lessons for Defenders: Governance Failures are Technical Vulnerabilities

This attack was not characterized by zero-day exploits or novel malware. It succeeded because basic governance and monitoring failures allowed attackers to operate undetected with full domain administrator privileges for more than a month. The absence of adequate identity controls, endpoint visibility, and privileged-access oversight created conditions in which a routine intrusion escalated into a full-scale ransomware event.

Reviewing the failure points in this case underscores a broader lesson: weaknesses in governance, visibility, and operational discipline are not abstract risks; they are concrete technical vulnerabilities. Addressing them requires treating identity, monitoring, and access controls as core security infrastructure, not secondary safeguards.

> “Ransomware has never been a single event to recover from. It is a sustained pressure campaign that exploits identity gaps, fragmented visibility, and slow decision-making, long before encryption ever appears. Incident response in 2026 is about containing business impact early, not negotiating after control is lost.”

TIM OTIS
Head of Incident Response Services



FROM RECON TO NARRATIVE CONTROL: CYBER'S OPERATIONAL IMPACT IN 2025 CONFLICTS

In 2025, cyber operations functioned as an integrated part of warfare along with air power, artillery, and special operations. The impact of cyber operations was realized through sustained interaction with military, political, and informational processes, instead of isolated technical effects.

The cyber operations we observed in 2025 served a small number of recurring functions, including Positioning and Conditioning Activity, which established and maintained access to key systems ahead of escalation; Operational Support Activity, which enabled or reinforced ongoing military, political, or influence operations; Direct Effect Activity, which caused immediate disruption, degradation, or denial; and Narrative Shaping Activity, which influenced information flows, public perception, and messaging during conflict.

These roles are not sequential and frequently overlap, as the same access, capability, or operation may serve different purposes over time. For example, the Russian-linked intrusion into Ukrainian power and telecom networks was used simultaneously for battlefield reconnaissance, disrupting civilian services during missile strikes, and sending a post-strike message to the victims that the hostilities were far from over.

In this section, we examine four active conflicts in 2025:

- Russia-Ukraine
- Iran-Israel
- India-Pakistan
- Thailand-Cambodia

Figure 1 - Components of major cyber functions in a military conflict

POSITIONING AND CONDITIONING ACTIVITY

Positioning and Conditioning activities played an essential role in most of the 2025 conflicts, shaping the technical and informational environment without producing immediate, visible effects.

Typical activities include establishing initial access and a long-term foothold, mapping infrastructure and dependencies, supply-chain reconnaissance, and conducting broad espionage activity. This stage also includes creating and introducing themes, personas, or information channels to influence audiences in later stages.

In fully developed cyber conflicts, positioning activity was persistent and cumulative over time, despite changes in tempo and shifts between kinetic (directly destructive warfare) and non-kinetic phases. In the Russian–Ukrainian conflict, this resulted in sustained access to logistics, transportation, and government-adjacent networks. Russian military intelligence-affiliated operators systematically targeted more than 10,000 internet-connected Ukrainian cameras positioned near roads, border routes, and infrastructure hubs. Access was obtained through exposed live camera feeds, weak credentials, and misconfigured devices, allowing operators to monitor movement around critical facilities during the first half of the year.

“ACCESS WAS OBTAINED THROUGH EXPOSED LIVE CAMERA FEEDS, WEAK CREDENTIALS, AND MISCONFIGURED DEVICES, ALLOWING OPERATORS TO MONITOR MOVEMENT AROUND CRITICAL FACILITIES DURING THE FIRST HALF OF THE YEAR.”

This activity was not only restricted to Ukrainian territory but also extended into Western logistics and supply-chain networks, expanding visibility into territories well beyond the immediate operations. The Russian-linked APT28 group leveraged existing, multi-year access to Western rail, maritime, and aviation logistics networks, as well as cloud platforms used to coordinate shipment routes supporting Ukraine.

With many Russian intelligence officers expelled from European Union states, Russian intelligence services reportedly relied on local intermediaries, including minors, to plant Wi-Fi sniffers, rogue access points, and signal-collection devices near embassies and government facilities. The goal was not to cause immediate disruption but to enable long-term situational awareness, dependency mapping, and options for later use.

Similar patterns were observed in Iranian-linked activity targeting Israeli civilian and commercial infrastructure, where access to cameras, IoT devices, and networked services was also for the purpose of conditioning and familiarization, and not immediate operational effects. Check Point Research (CPR) recorded more than a 1,200 percent surge in exploitation attempts, with some targeting outdated devices and vulnerabilities of

specific camera vendors, combined with weak passwords, to pull live camera feeds throughout the country.

Iranian state-affiliated hacktivist groups such as Handala and CyberAv3ngers also conducted reconnaissance on Israeli industrial control systems, operational technology, and satellite-linked infrastructure, scanning for exposed entry points that could be leveraged during escalation. These probes underscored a widening interest in Israel's physical infrastructure and support systems. These activities together illustrate a layered reconnaissance strategy: visual surveillance from hijacked cameras, credential-based access to institutional systems, penetration of IT suppliers, and probing Israel's industrial and satellite assets. Iran's pre-strike digital preparation was broad, persistent, and distributed across multiple sectors.

In other conflicts, positioning activity lacked this level of sophistication and was often short-lived. While there were attempts to gain access and conduct reconnaissance, they often lacked persistence and depth.

Reconnaissance helped shape the early phase of the Indian–Pakistani conflict. Following the April 2025 terrorist attack in Pahalgam, for which India publicly blamed Pakistan, the Pakistani-linked APT36 group deployed phishing lures disguised as incident-related reports to compromise Indian defense personnel. The malicious documents delivered Crimson RAT, enabling credential theft, persistent access to sensitive accounts, and observation of internal defense workflows.

Positioning and conditioning activities also played a prominent role in the conflict between Thailand and Cambodia. After a border clash in May 2025, Thai and Cambodian-affiliated groups probed each other's government platforms, public-sector web services, and communication channels.

While less sophisticated or limited in scope than the other conflicts, both the Indian-Pakistani and Thai-Cambodian conflicts illustrate the same emphasis on positioning for potential escalation and expanding future options instead of achieving immediate results.

Operational Support Activity

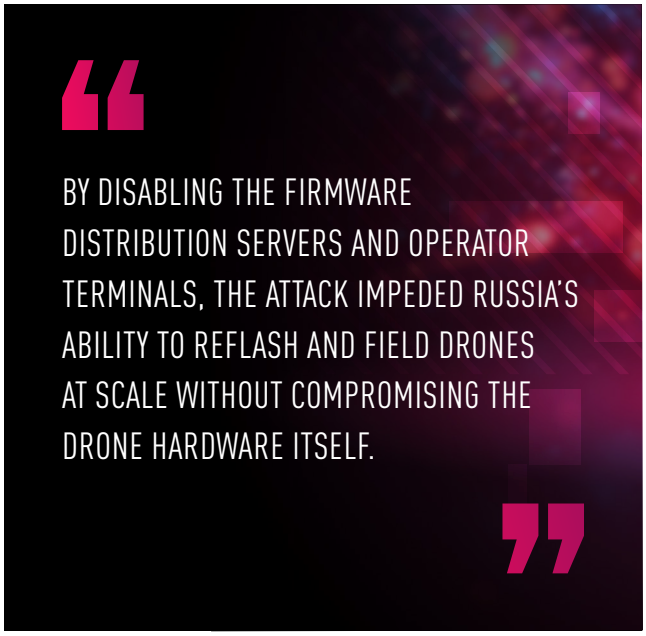
In 2025, operational support became more prominent as cyber operations were increasingly part of ongoing physical events instead of being conducted in isolation. These typically time-sensitive actions enabled, amplified, or synchronized activity in other domains and were closely aligned with unfolding military or political developments.

Support activity includes collecting targeted intelligence, monitoring logistics and movement in real-time, disrupting communications, and applying psychological pressure aligned with kinetic operations. Unlike in the earlier positioning stage, the espionage is narrower in scope, more targeted, and aimed at achieving immediate tactical results. Logistics intelligence shifted from static mapping to tracking, prioritization, and real-time situational awareness.

In the Israeli–Iranian confrontation, previously compromised civilian surveillance infrastructure was activated to provide operational visibility. In June 2025, Iranian operators accessed security cameras surrounding the Weizmann Institute and adjacent street-facing systems, gaining live feeds

of roads, parking areas, and movement patterns. These feeds were monitored in the hours leading up to and during an Iranian missile strike on the Weizmann Institute, repurposing consumer-grade sensors into an improvised reconnaissance network supporting real-world targeting.

Researchers also observed operational support activity when a Ukrainian-linked cyber operation disrupted Russian battlefield drone operations by targeting the infrastructure used to distribute and deploy custom firmware for modified civilian UAVs. By disabling the firmware distribution servers and operator terminals, the attack impeded Russia's ability to reflash and field drones at scale without compromising the drone hardware itself.



Russia's cyber operations showed a similar close integration with kinetic activity. The Russian-linked APT44 group (also tracked as Sandworm) frequently operated in parallel with missile and drone barrages, launching cyber attacks against logistics, agriculture, and energy networks to disrupt Ukrainian infrastructure and complicate service restoration. Analysis noted that major missile strikes were often followed by coordinated cyber activity and subsequent surges of pro-Russian hacktivist DDoS attacks, as well as Telegram-controlled narratives amplifying their successes.

During the May 2025 confrontation along the Line of Control, Pakistani cyber operations unfolded in parallel with drone and missile exchanges. Indian authorities reported large-scale cyber activity, including DDoS attacks, malware intrusions, and GPS spoofing, together with periods of kinetic escalation. While attribution and scale varied across the various reports, the timing of these cyber attacks reinforces assessments that cyber disruption functioned as a complementary tool for degrading battlefield awareness and complicating decision-making during high-intensity exchanges.

Cyber activity tracked closely with physical developments in more localized conflicts as well. Following a May 2025 border clash, Cambodian-aligned hacktivist groups rapidly escalated attacks against the Thai government, military, and civilian networks. Activity surged again immediately after a televised Thai military announcement signaling heightened readiness.

These cases illustrate how, in 2025, by laying the groundwork in an earlier stage, synchronized cyber operations increasingly functioned as real-time enablers of military and political action.

Direct Effect Activity

Direct effect activity includes cyber operations intended to produce immediate and visible impact. These actions target systems, data, or services directly, with results that can be measured in technical, economic, or operational impacts.

In 2025, direct cyber effects were used selectively. Destructive attacks, ransomware campaigns, and data leak operations garnered significant attention, but they frequently only played a supporting role in the broader conflict. Disruptive actions against financial institutions, government services, and civilian resilience sectors tend to be limited in duration yet still result in major damage, as they affect critical infrastructure, degrade function, and force rapid recovery under pressure.

In the Israeli–Iranian conflict, Iranian state-linked operators and affiliated hacktivist ecosystems increasingly focused their attention on civilian sectors such as healthcare, research institutions, and financial services. Iranian-linked actors repeatedly attempted to compromise hospital networks, exfiltrate sensitive medical data, and interfere with clinical operations. These incidents were not treated as isolated intrusions but as part of a broader pattern of coercive disruption aimed at undermining essential services and public confidence.

Direct effect activities were also used against Iran, where disruptive cyber operations targeted institutions central to the country's financial well-being. A group operating under the name Predatory Sparrow conducted a pair of high-impact attacks, first against Bank Sepah, resulting in widespread service outages and reported destruction of core banking data, and later against Nobitex, Iran's largest cryptocurrency exchange, where digital assets were rendered inaccessible and proprietary source code was publicly leaked. While the operation's sponsorship and strategic direction were not independently confirmed, the attacks demonstrated that critical Iranian institutions could be disrupted rapidly and at scale.

Iranian authorities responded by sharply restricting nationwide internet access for more than a day, a "defensive" measure aimed at reducing further intrusion attempts. The restrictions resulted in immediate hardship to the civilian population, disrupting access to banking, news, and basic communications during a period of heightened uncertainty.

Russia's cyber activity in Ukraine showed a deliberate pattern of destructive intrusions timed to coincide with physical strikes. The Russian-linked APT44 group deployed wiper malware against government, energy, logistics, and agricultural sectors in an attempt to weaken Ukraine's economic resilience. The group frequently operated in parallel with missile and drone barrages, which not only amplified the disruption to the networks but also complicated Ukrainian attempts to restore services.

While many such operations are visible by design, their defining feature is that the functional impact itself carries value, even in the absence of public attribution or sustained attention.

Narrative Shaping Activity

In 2025, Narrative Shaping activity was central to cyber operations, and, in many cases, the results were more enduring than technical disruptions. The purpose of this activity is to shape perception, signal capability or intent, and influence domestic or international audiences. Visibility, attribution, and interpretation are therefore central to their impact.

Influence operations, hack-and-leak campaigns, defacements, and public claims of responsibility were prominent in multiple conflicts. Russia maintained its long-standing emphasis on influence operations, with coordinated narratives amplified through cyber-enabled channels, proxy outlets, and automated content generation. Technical damage was frequently employed as a signaling mechanism, and the narrative impact was more important than the scale of disruption.

One example of this is a ransomware attack against the Shamir Medical Center in Israel. The intrusion initially appeared as a conventional financially motivated ransomware incident and leveraged Qilin, a ransomware-as-a-service platform typically associated with profit-driven criminal activity. Subsequent investigation linked the operation to Iranian state-aligned actors. Following public exposure, Qilin withdrew its ransom demands and removed the hospital from its list of victims.

As missile exchanges intensified in June 2025, Iranian information operations focused on destabilizing the civilian population by eroding trust in Israel's emergency-response systems. They released a wave of fake Home Front Command alerts, crafted to appear indistinguishable from official rocket-warning notifications, during periods of heightened threat perception. Fraudulent SMS messages warning of fabricated terror attacks, resource shortages, and infrastructure failures circulated alongside AI-generated imagery and coordinated hashtag campaigns portraying Israeli society as collapsing under sustained military pressure. None of this was true. Israeli authorities reported more than 1,200 coordinated social-engineering operations targeting the public during this period.

During the Thai-Cambodian border tensions, cyber operations similarly prioritized destabilization over tactical effects. The Thai government's platforms and media infrastructure absorbed more than 223 million malicious requests within a single 24-hour period, overwhelming public-facing services immediately following military and political signaling. Cambodia responded with public accusations of Thai-linked intrusions across multiple ministries, while the Cambodian-aligned hacktivist group KH Nightmare leaked approximately 800GB of alleged government data, amplifying uncertainty and eroding confidence in the government.

Although these incidents involved large-scale disruption and data exposure, their significance depended on visibility, attribution, and interpretation rather than specific technical damage. Their primary effect was to strain administrative confidence and shape decision-making under stress, illustrating how technical disruption can function primarily as a signaling mechanism in contemporary conflict.

These activities all illustrate efforts to exert psychological pressure. The purpose of the operations was not to inflict physical harm, but to erode trust in official communication channels. In several cases, the most influential cyber operations were not those that disabled infrastructure, but those that caused civilians to question alerts, doubt warnings, and experience sustained uncertainty or anxiety.

Similar dynamics were evident in Russian-affiliated information operations during 2025. Those operations increasingly focused on dominating the hours immediately following offensive missile strikes or cyber incidents, deploying rapid, high-

volume messaging campaigns intended to outpace verification and correction. Analysts observed coordinated surges of parallel narratives across hundreds of channels, shaping perception before official information could be released.

A notable development was the expansion of AI-driven content saturation. The pro-Kremlin Pravda network evolved into one of the world's largest disinformation engines, publishing up to 23,000 articles per day across hundreds of sites, including extensive English-language items that increased visibility in search engine results. Experts warned that the sheer volume enabled forms of so-called "LLM grooming," in which large language models are continuously exposed to skewed narrative inputs.

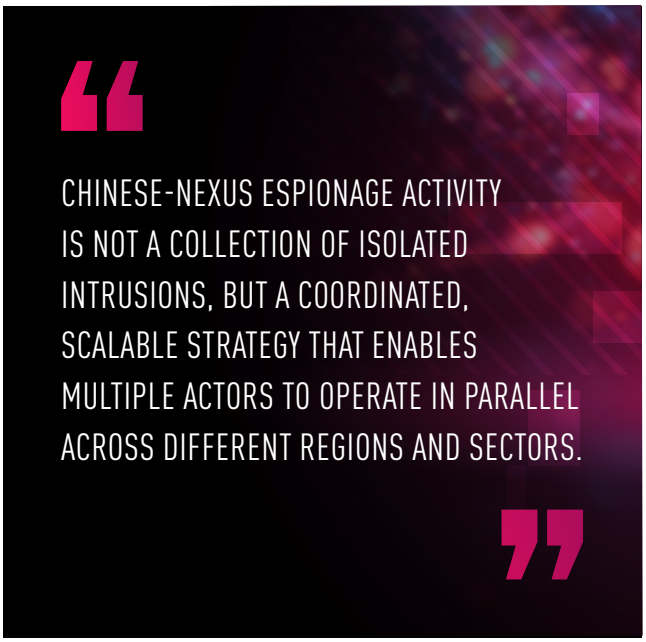
Russian cyber campaigns impersonated European media outlets, manipulated public debate around military aid to Ukraine, and targeted elections in Germany, Romania, and Moldova, with the intent to erode democratic confidence and institutional trust. These patterns illustrate a post-strike psychological doctrine in which narrative floods and AI-scaled influence operations are used to saturate perception at moments when populations are most vulnerable to fear, uncertainty, and misinformation.

The conflicts of 2025 illustrate that cyber activity has matured into a persistent and integrated component of modern warfare rather than a discrete or exceptional instrument. It proved invaluable in multiple conflicts, shaping environments before escalation, enabling operations as events unfolded, imposing friction through targeted disruption, and exerting sustained psychological pressure long after the kinetic effects were felt. The practical value of cyber operations in 2025 lies in the ability to compound other attack venues, exploit uncertainty, and operate continuously below traditional thresholds of escalation. Understanding cyber activity through its functional roles demonstrates why its cumulative impact is increasingly shaping our conduct and perception of war.

THE DOMINANCE OF CHINESE-NEXUS CYBER THREATS

Cyber operations in 2025 no longer respect national borders. Chinese-affiliated threat actors run concurrent campaigns across multiple continents and critical sectors. Actors treat global telecommunication organizations, cloud service providers, enterprise infrastructure, and the surrounding internet ecosystem as a shared operational environment, effectively creating a single attack surface.

The attackers' approach leverages mature access platforms, such as ShadowPad and PlugX, and higher-end capabilities, including BRICKSTORM, malware designed for stealth and long-term persistence across edge devices and virtualization infrastructure. Services that are exposed to the internet and systems that mediate identity, traffic, and trust are consistently leveraged across targets. Chinese-nexus espionage activity is not a collection of isolated intrusions, but a coordinated, scalable strategy that enables multiple actors to operate in parallel across different regions and sectors.



CISA's advisory. "Countering Chinese State-Sponsored Actors Compromise of Networks Worldwide to Feed a Global Espionage System,"details Salt Typhoon's activity targeting networks across the globe, including telecommunications, government, transportation, and military infrastructure. While threat actors focus on large backbone routers of major telecommunications providers, as well as provider edge and customer edge routers, they also leverage compromised devices and trusted connections to pivot into other networks. Notably, this campaign underscores that global access does not always require powerful new tools. Durable positioning within widely deployed infrastructure and the trust relationships surrounding it can be enough to create international reach.

The advisory was co-signed by 23 authoring and partner agencies from the United States and international partners across Europe, Oceania, and Asia. This broad coordination reflects the global scope of the issue and the need for joint defensive action around widely deployed infrastructure.

Unmonitored Devices Exploitation

Salt Typhoon represents only one element of a far wider set of perimeter-focused operations. As discussed in the "Unmonitored Devices: The Attackers' Launch Base" chapter, throughout 2025, the exploitation of edge infrastructure remained central to enabling lateral movement and long-term access, particularly when patch cycles, visibility, and detection controls lag behind endpoint threats.

Threat actor UNC5221 has historically focused on edge devices and continued this activity throughout 2025, particularly targeting Ivanti Secure VPN solutions. The actor exploited zero-day and recently disclosed vulnerabilities to deploy custom malware implants in multiple incidents. These operations exploited Ivanti Secure VPN zero-day vulnerabilities, including CVE-2025-0282 and CVE-2025-22457, to deploy custom platform-specific SPAWN malware. Together, these implants provide capabilities ranging from persistent access and traffic tunneling to log tempering, enabling stealthy, privileged access on the compromised appliance.

BRICKSTORM, also attributed to UNC5221, achieved greater prominence following the F5 Networks compromise disclosure, revealing a disproportionate risk to software vendors. By transforming edge infrastructure into covert environments for intellectual property theft and downstream compromise, threat actors can convert vendor-side access into exploitation blueprints that can be leveraged against their customers at scale.

Other threat actors with advanced technical skills were also observed targeting perimeter infrastructure. UAT4356, which was responsible for the ArcaneDoor campaign, targeted Cisco ASA 5500-X firewalls to deploy the Ray Initiator bootkit, and UNC3886 deployed custom implants on Juniper routers. Collectively, these campaigns illustrate a broader strategy among Chinese-nexus espionage actors: a sustained investment in zero-day edge exploitation and the development of platform-specific tooling to gain access to vulnerable and unmonitored organizational systems.

“THESE CAMPAIGNS ILLUSTRATE A BROADER STRATEGY AMONG CHINESE-NEXUS ESPIONAGE ACTORS: A SUSTAINED INVESTMENT IN ZERO-DAY EDGE EXPLOITATION AND THE DEVELOPMENT OF PLATFORM-SPECIFIC TOOLING TO GAIN ACCESS TO VULNERABLE AND UNMONITORED ORGANIZATIONAL SYSTEMS.”

Industrialized Tradecraft and a Shared Malware Ecosystem

Chinese-nexus cyber operations employ a deliberate, well-resourced campaign strategy, using shared tooling, repeatable operational playbooks, and consistent execution across targets and regions.

In multiple Chinese-nexus operations worldwide, activity frequently follows a recognizable modus operandi. Throughout 2025, Check Point Research observed multiple campaigns on nearly all continents in which initial access and execution commonly leveraged DLL side-loading, a technique allowing malicious code to run under the cover of legitimate, trusted software. This is commonly paired with staged loaders and modular backdoor ecosystems, where malware families such as PlugX and ShadowPad are widely seen in different clusters and intrusion sets, evidence of a shared tooling ecosystem. These backdoors serve as operational hinges for command execution, credential access, and reconnaissance.

Post-compromise, operators frequently expand their control using “living-off-the-land” techniques and administrative protocols that blend into the system’s own routine IT activity, such as Remote Desktop Protocol for lateral movement and hands-on operations. To maintain persistence and operational flexibility, threat actors may also deploy VPN software under modified names, a recurring technique in Chinese-nexus playbooks that embeds remote control in seemingly normal connectivity. What appears to be an ordinary sequence of technical actions is, in fact, an exercise in trust building designed to avoid suspicion.

The post-compromise deployment of ShadowPad is a recurring feature in our investigations of campaigns across Europe, Africa, Central Asia, and the APAC. Its modular architecture supports data exfiltration, remote command execution, lateral movement, and credential harvesting, which allows flexible deployment as mission requirements change. ShadowPad's repeated appearance in otherwise unrelated events demonstrates that stable, adaptable tooling is effective across diverse environments, reflecting a calculated approach to scalable, globally distributed data collection.

In recent InkDragon campaigns targeting Europe, Southeast Asia, Africa, and South America, ShadowPad was the core mechanism for persistence and execution. By deploying a custom ShadowPad IIS Listener module, the actor also reused the compromised infrastructure as a hub for further operations. This technique reflects a broader trend of turning footholds into platforms, and platforms into supply chains.

Exploitation of Trusted Enterprise Infrastructure

Throughout 2025, exploiting Microsoft internet-facing servers was a prominent intrusion vector, and helped threat actors to weaponize newly disclosed vulnerabilities rapidly. This strategy enables scalable access before vulnerabilities are patched, highlighting the use of ubiquitous enterprise platforms as long-term access vectors rather than one-off intrusion opportunities.

ToolShell is a particularly concerning exploit chain that enables unauthenticated remote code execution against on-premise internet-facing Microsoft SharePoint servers. Although publicly disclosed on July 19, the exploitation was first observed earlier in the month as a zero-day exploit against a small number of global government organizations. The intrusions were typically followed by targeted data theft using custom web shells

"THROUGHOUT 2025, EXPLOITING MICROSOFT INTERNET-FACING SERVERS WAS A PROMINENT INTRUSION VECTOR, AND HELPED THREAT ACTORS TO WEAPONIZE NEWLY DISCLOSED VULNERABILITIES RAPIDLY."

designed to extract cryptographic IIS/ASP.NET machine keys from compromised environments. In recent InkDragon intrusions, initial access was gained using ToolShell. For New post-exploitation phase, to facilitate lateral movement and data exfiltration, the actor deployed ShadowPad or FinalDraft malware.

The APT group, Rude Panda, abuses misconfigurations rather than vulnerabilities. Its campaigns conducted throughout late 2024 and into 2025 focused on compromising Microsoft IIS servers using publicly available, static ASP.NET machine keys. Following initial access, the attackers deployed a custom malicious IIS module named "HijackServer." This activity resulted in the compromise of hundreds of servers globally, which were subsequently leveraged to support fraudulent operations, including search engine optimization (SEO) manipulation and cryptocurrency schemes.

One of the clearest examples of the rapid weaponization cycle is CVE-2025-59287, a remote code execution vulnerability in Windows Server Update Services (WSUS). The vulnerability was weaponized within days publishing a proof-of-concept (PoC). Similar to previous campaigns, the post-compromise payload was ShadowPad and deployed via DLL side-loading. By abusing WSUS, the attackers gained high-privilege execution inside the Windows update infrastructure, on which most networks depend.

These cases demonstrate that the central role of IIS and WSUS amplifies risk by design. When components responsible for distributing trust are subverted, the impact is broad and difficult to contain.

Figure 1: Significant Chinese Affiliated Threat Activity in 2025

What Will Change in 2026: Deny Persistence, Not Just Intrusion

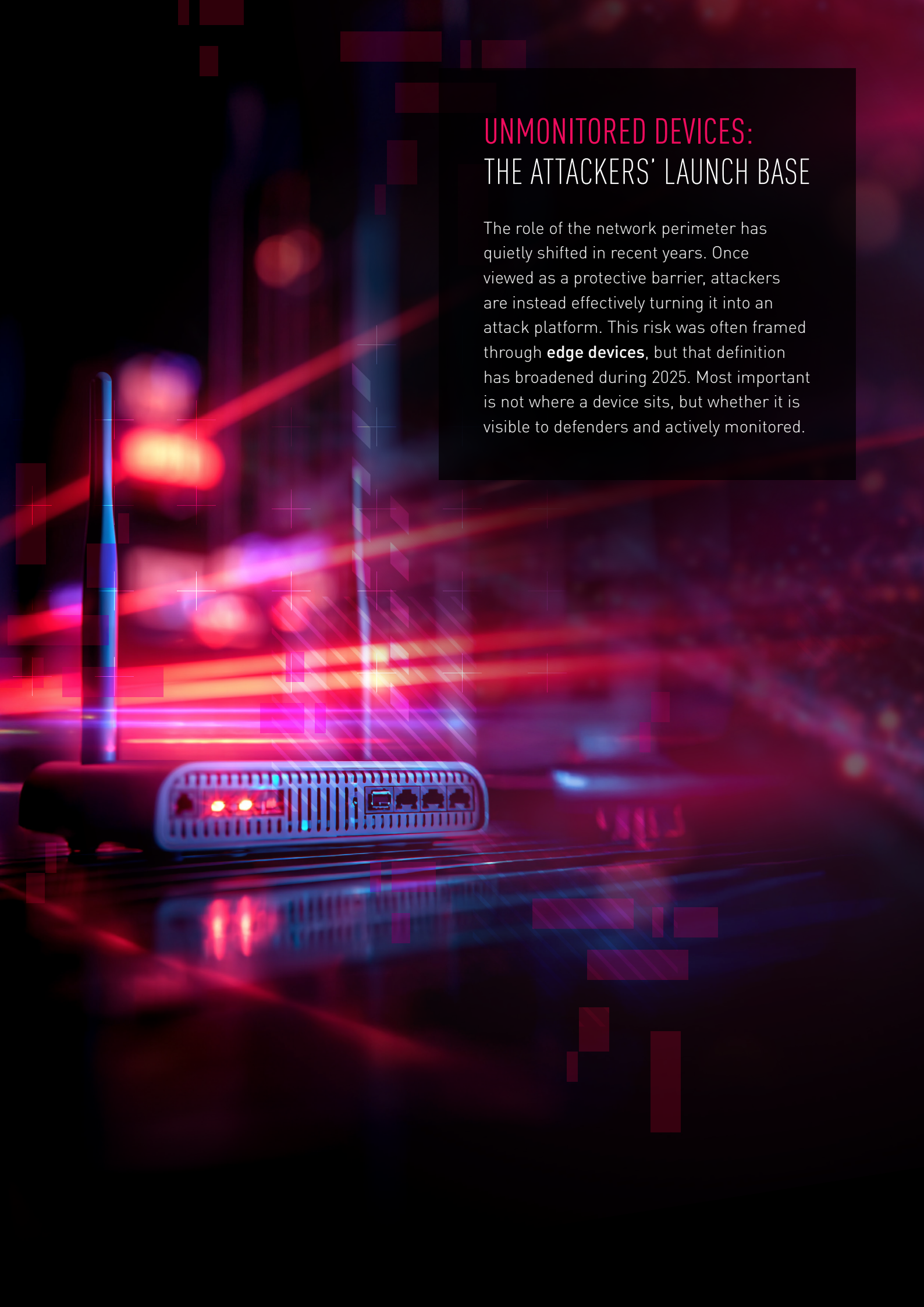
In 2025, our investigations showed a sustained increase in activity attributed to Chinese-nexus threat actors, with a global footprint across critical sectors. These campaigns are linked not only by shared tooling but also by consistent operational models, including edge-focused intrusion paths, exploitation of zero-day and one-day vulnerabilities, and an emphasis on stealth and long-term persistence.

The operations rely on a common set of malware families and techniques that enable data exfiltration, credential harvesting, and lateral movement, supported by C2 channels designed to be indistinguishable from modern enterprise traffic. A recurring theme is the abuse of trusted enterprise services, such as IIS and WSUS, to support stealthy persistence and operational scalability.

These campaigns follow a deliberate strategy of reusing proven tactics across regions and industries, turning ubiquitous enterprise platforms into durable access points. In 2026, the challenge extends beyond preventing initial compromises to stopping attackers from leveraging access to compromise downstream enterprises and maintain persistence within trusted services.

> This chapter makes clear that geopolitical cyber activity is no longer episodic or symbolic. It is persistent, coordinated, and directly tied to national objectives, requiring security leaders to treat nation-state threats as a standing operational risk rather than a special case.

ELI SMADJA
Security Research Group Manager



UNMONITORED DEVICES: THE ATTACKERS' LAUNCH BASE

The role of the network perimeter has quietly shifted in recent years. Once viewed as a protective barrier, attackers are instead effectively turning it into an attack platform. This risk was often framed through **edge devices**, but that definition has broadened during 2025. Most important is not where a device sits, but whether it is visible to defenders and actively monitored.

Unmonitored devices are particularly attractive targets due to their low visibility, privileged proximity to traffic and identity, and positioning within operational environments that are difficult to patch without downtime. Unmonitored devices, such as routers, firewalls, VPN appliances, and virtualization solutions, often operate in less monitored zones without Endpoint Detection and Response (EDR) capabilities, resulting in sparse logs and short retention windows.

Now attackers are using unmonitored devices as launching pads, easily blending into legitimate network traffic and can harvest credentials and move laterally before defenders can respond. In 2025, Chinese-affiliated operators extended this approach to vendors of these devices, targeting them not only for access but for internal knowledge, source code, and undisclosed vulnerabilities, ultimately preying on high-profile customers who trust those vendors to secure their environments.

UNMONITORED DEVICES ARE NOW USED AS ATTACKER LAUNCHING PADS

2025's defining pattern was persistence, with the exposure of long-running campaigns targeting unmonitored devices. As defenses in standard, well-monitored environments continue to improve, more attackers are choosing to shift their operations into unmonitored environments, where visibility is weaker and response times are slower. While this effort has been spearheaded by Chinese-affiliated actors, it is also becoming increasingly common among other state-linked operators, as well as cyber criminals. When attackers operate from unmonitored footholds, every signal becomes suspect, and every external dependency becomes a potential threat.

Zero-Day Exploitation and a Growing Custom Malware Ecosystem

Among the many campaigns targeting unmonitored devices, intrusions attributed to the Chinese-affiliated espionage actor UNC5221 stand out for their scale, efficiency, and longevity, showing how low-visibility infrastructure can serve as a quiet launch base for sustained operations. Since at least 2022, UNC5221 has been repeatedly exploiting zero-day vulnerabilities in Ivanti Connect Secure appliances, deploying BRICKSTORM malware.

BRICKSTORM is a custom implant designed to run inside multiple types of network appliances. In several cases, it was initially deployed on an edge device and then used to pivot deeper into internal unmonitored infrastructure, including VMware vCenter and ESXi. The malware is designed to blend in with typical operations and evade standard incident-response cycles, often persisting beyond remediation efforts. By operating from unmonitored systems with access to core network flows, BRICKSTORM enabled its operators to capture credentials and expand access into cloud environments. This progression illustrates how a single compromise in an unmonitored device can quickly escalate into broad identity and data exposure.

BRICKSTORM activity also demonstrates an alarming development: attackers now target the vendors that build these devices. In October, F5 Networks disclosed long-term unauthorized access to its internal BIG-IP development environment, linked to BRICKSTORM actors. During the intrusion, attackers exfiltrated source code, knowledge base content, and information on undisclosed vulnerabilities. These assets could be leveraged to develop reliable exploit paths against a broad customer base. When a development environment is compromised, the security posture of the entire product ecosystem is also impacted. Malware persistence on the

vendor side can easily become customer-side exposure, accelerate exploit development, and enable broader compromise across sectors that depend on those products.

Multiple other China-affiliated actors are weaponizing unmonitored network appliances with custom, platform-device-specific implants. The U.K. National Cyber Security Centre (NCSC) issued a public advisory on a new implant campaign targeting Cisco ASA-5500-X firewalls, which are end-of-life and no longer supported. The activity, tracked as UAT4356 and attributed to a China nexus, involved active exploitation of CVE-2025-20333 and CVE-2025-20362 on compromised devices to deploy a previously undocumented toolkit.

VENDOR-SIDE MALWARE PERSISTENCE DRIVES CUSTOMER EXPOSURE AND BROADER COMPROMISE.

Once these unmonitored appliances were compromised, attackers gained high-value access to network traffic, credentials, and administrative functions. Designed to survive reboots and, in some cases, even firmware upgrades, the malware enabled attacker control, including bypassing AAA (Authentication, Authorization, and Accounting) checks, executing commands, conducting covert packet capture, and exfiltrating sensitive data.

Another China-nexus actor, tracked as UNC3886, was observed exploiting CVE-2025-21590, a vulnerability in Junos OS, Juniper Networks' operating system for routing, switching, and security devices. UNC3886 abused this flaw to run malware from within legitimate Junos OS system processes, enabling stealthy persistence and post-exploitation activity without breaking file integrity checks or triggering standard security controls, making it particularly difficult to detect unmonitored Juniper appliances.

UNC3886 historically focuses on maintaining long-term access to victim environments and has previously been observed exploiting zero-day vulnerabilities to deploy custom malware on Fortinet devices. In intrusions involving Juniper routers, operators deployed multiple implants in parallel, with core capabilities such as remote file upload and download, interactive shell access, and connection relay. The implants operate within Junos' underlying FreeBSD environment and are designed to masquerade as legitimate system processes. In some cases, they even interact directly with legitimate Junos OS daemons, including patching their memory at runtime to suppress logging and telemetry.

These operations reflect a high degree of research effort and sophistication, with an emphasis on stealth and maintaining persistent, long-term, and covert access, particularly to devices that are not continuously monitored.

Misconfiguration, Not Zero-Day

A different approach was observed in a long-running Russian state-sponsored campaign attributed to the Sandworm group that targets Western critical infrastructure. Unlike the Chinese actors, which relied on direct compromise of devices, this activity primarily abused misconfigured network devices, gaining access through exposed management interfaces rather than exploiting vulnerabilities in the platforms themselves.

The attackers gained administrative access and then leveraged native traffic-capture and monitoring capabilities to intercept network traffic passively. This enabled them to harvest credentials, session cookies, and authentication tokens, which were subsequently used to access

legitimate organizational services such as VPNs, identity providers, and cloud management consoles. Establishing persistence and lateral movement using valid credentials eliminates the need for custom malware implants, while making the activity harder to separate from everyday administrative use, especially when device access and telemetry are not consistently monitored.

Regardless of whether a campaign relies on exploitation or misconfiguration, unmonitored devices have become a critical and increasingly attractive attack surface for sophisticated threat actors, providing a low-friction path to stealthy access, credential theft, and long-term operational persistence.

Ransomware: N-Day Exploitation at the Edge

Targeting unmonitored devices is not new for cyber criminals. Ransomware operators have targeted VMware ESXi and vCenter environments for years because they sit at the center of server operations yet offer minimal visibility. This targeting does not stand alone, as financially motivated actors have increasingly adopted techniques long associated with state-sponsored operators, including the systematic exploitation of patched-but-still-exposed vulnerabilities, known as n-days, in network appliances.

One of the most notable examples of financially motivated attackers using unmonitored devices as attack points, by weaponizing n-day vulnerabilities, involved Akira ransomware. In mid-2025, an increase in Akira intrusions targeting SonicWall appliances was observed. Multiple incidents involved unauthorized access through SonicWall SSL VPN services, followed by the deployment of ransomware, often within hours of the initial compromise. Victims spanned multiple sectors and organizations regardless of size, indicating opportunistic mass exploitation rather than targeted intrusion.

While a zero-day vulnerability in SonicWall SSL VPN was initially suspected, this spike was largely tied to ongoing abuse of CVE-2024-40766. This year-old access control vulnerability allows local user passwords carried over during migrations to remain valid. As a result, credentials harvested when devices were vulnerable can later be reused by threat actors, even after the affected devices have been patched. This illustrates the long-tail risk of credential exposure stemming from vulnerabilities in unmonitored devices.

Qilin, the most prolific ransomware operation observed in 2025, similarly relied on exploiting unmonitored network appliances for initial access. The group was observed exploiting CVE-2024-21762 and CVE-2024-55591, which affected FortiOS and FortiProxy SSL VPN devices. The first vulnerability enables remote execution of arbitrary code or commands, while the second allows attackers to obtain super-administrator privileges. Together, these flaws provided rapid, privileged access to victim environments through commonly deployed perimeter infrastructure.

This pattern was not limited to established ransomware brands. In early 2025, vulnerabilities in Fortinet devices were exploited by a newly identified actor tracked as Mora_001. The group exploited CVE-2024-55591 and CVE-2025-24472, both of which allow unauthenticated attackers to gain super-admin privileges on vulnerable FortiOS devices. Following initial access, the actors mapped internal networks, moved laterally using newly created VPN accounts, and ultimately deployed a ransomware strain called SuperBlack, built using a leaked LockBit builder.

Together, these incidents highlight how ransomware operators favor n-day exploitation of unmonitored devices to achieve fast, privileged access. This has become a key complement to their long-standing focus on ESXi and vCenter: unmonitored devices provide the entry point, and virtualization platforms offer the scale and leverage. By operating from systems that sit outside standard endpoint visibility, attackers shorten the time from entry to impact, bypass common controls, and convert external exposure directly into rapid financial harm.

Looking Ahead

Unmonitored devices continued to evolve this year from operational network systems into platforms for malware activity. Limited visibility and inconsistent monitoring make them well-suited for long-term access, and their role in handling high-value traffic and authentication flows enables attackers to expand access well beyond the initial point of compromise. At the same time, the targeting and compromises within vendor environments increase the risk that intrusions translate into downstream exposure for customers.

To mitigate the effects of long-running intrusions involving unmonitored devices, end-of-life systems must be retired. For devices that are still supported, vendors must provide stronger monitoring capabilities, richer forensic artifacts, and more resilient security controls by default. Without these changes, compromises in low-visibility infrastructure will remain challenging to detect and even harder to contain.

> One of the most dangerous attack surfaces is infrastructure we assume is trusted. Edge devices operating with limited visibility allow attackers to establish persistent footholds rather than one-time entry points.

ALEXANDRA GOFMAN

Technology Leader
Threat Intelligence

03

AI LANDSCAPE IN CYBER SECURITY

AI LANDSCAPE: FROM INTEGRATION TO AUTONOMY

In 2025, Artificial intelligence (AI) was so deeply embedded in cyber activity that distinguishing "AI-related attacks" from general digital operations became increasingly challenging. In contrast to 2023–2024, when attackers' use of AI was easily recognizable, in 2025 AI use became so commonplace that it faded into the background of attack operations. AI now underpins software development, social engineering malware design, data mining, influence operations, reconnaissance, vulnerability discovery, and even post-exploitation activity.

AI is now used everywhere yet remains rarely visible. Most malicious outputs seldom reveal if AI contributed to their creation or execution. Our April 2025 State of AI in Cyber Security report warned that as AI models become integrated into daily work, the boundary between "AI-enabled" and conventional threats would blur. By the end of 2025, that prediction will have come to pass.

Throughout 2025, threat actors not only refined and expanded their use of AI but also increasingly attempted to target the AI ecosystem itself. As enterprises adopt agentic frameworks, MCP servers, and locally deployed models, these environments have become the new attack surfaces.

The following chapter examines AI's dual role in today's threat landscape. First, we outline the growing class of attacks on AI services and agentic systems, where misconfigurations, prompt manipulation, and vulnerabilities in AI-connected tools create opportunities for exploitation. Second, we assess attacks enabled by AI, including identity theft and impersonation, AI-assisted malware development, automated reconnaissance, and the broader optimization of AI across criminal and state-sponsored activity. Finally, we determine what changed in 2025 and the implications for 2026.

Our focus is on real-world, in-the-wild evidence collected throughout 2025, including attacker operations, underground services and discussions, published incidents, and law enforcement findings.

AI SERVICES AS AN ATTACK SURFACE

As AI tools and services become fully integrated into every aspect of daily corporate life, their access to data, context, and downstream systems is increasing astoundingly fast. AI assistants and agents are involved in processing emails, documents, calendars, web content, and internal knowledge databases. As a result, AI is becoming an increasingly attractive attack surface.

Direct and Indirect Prompt Injection Attacks

A clear manifestation of this trend is the rise of direct and indirect prompt injection attacks. Attackers continued to turn prompt injection into a pervasive threat affecting both direct model interactions and autonomous agent workflows. Data from Lakera, a Check Point Company, shows that direct LLM manipulation is achieved through role-play setups, hypothetical scenarios, and obfuscation tricks. In such attacks, client-facing LLM-based services are targeted to expose restricted information. In indirect prompt injection attacks, malicious instructions are embedded within otherwise legitimate content that AI systems access during everyday workflows.

One reported research example involved malicious Google Calendar invitations that contained hidden instructions inside event descriptions. When processed by Google's Gemini assistant, the injected content influenced downstream behavior, enabling unauthorized actions such as sending messages, accessing application context, and interacting with connected smart home devices. The attack was possible due to the AI assistant's trusted access to calendar data and integrated services.

During the year, Google also issued a global advisory after observing invisible HTML-based injections that could manipulate AI summarization features within Gmail, demonstrating how subtle these attacks are and how difficult they are to detect. Meanwhile, Check Point Research documented real malware samples embedding natural-language instructions designed to mislead AI-powered detection tools, signaling that attackers are attempting to bypass LLM defenses.

Similar risks were observed in enterprise environments. Research demonstrated how AI systems integrated into corporate workflows could be manipulated through documents, tickets, or shared content containing concealed instructions. When AI assistants summarized or processed this material, the embedded payloads altered the model's behavior, leading to the unintended disclosure of sensitive information or unsafe tool execution. These findings underscore that indirect injection turns everyday business artifacts into potential execution vectors once they are implicitly trusted by AI-driven automation.

In 2025, Check Point Research disclosed a command injection vulnerability in OpenAI's Codex CLI, an AI-powered coding assistant designed to execute commands on a developer's local machine. The flaw allowed untrusted input to influence command execution, effectively enabling arbitrary commands to run in the host environment. The case illustrates how, once agentic AI tools are granted execution privileges, they can turn input manipulation into direct system compromise, even outside formal agent frameworks.

Lakera, a Check Point Company, provided additional validation for the risk of in-direct injection. In its Q4 2025 review of real-world agent-related attacks, it observed that indirect prompt injection attempts often proved more effective than direct. The report documented multiple cases in which hidden instructions embedded in emails, documents, or web pages influenced agent behavior, resulting in unintended tool invocation, leaking sensitive data, or suppressing safety constraints. Notably, The report found that these attacks frequently required fewer attempts than direct prompt injection as they exploited the agent's normal operating assumptions instead of trying to override safeguards.

Model Context Protocol (MCP) Under Attack

Model Context Protocol (MCP), the mechanism that allows LLMs to invoke external tools, is currently one of the most lucrative parts for attackers in the AI attack surface. Throughout 2025, Check Point Research and other researchers exposed structural weaknesses across the MCP ecosystem, which include servers, tool configuration files, IDE integrations such as Cursor and VS Code plugins, and the broader community of third-party nodes. Check Point Research published an RCE vulnerability in Cursor's implementation, known as MCPoison, which stemmed from the IDE's implicit trust in modified MCP configuration files. Other researchers reached similar conclusions in a separate investigation, finding that a large majority of publicly exposed MCP servers leaked sensitive information such as API keys, making them easy to compromise.

A recent review by Lakera found security vulnerabilities in 40% of the approximately 10,000 MCP servers that were probed. Seven percent were vulnerable to "Path Traversal" attacks, and Lakera found at least one secret API key in 8% of the servers. Two percent were vulnerable to SQL injection attacks, and 6% were vulnerable to Command or Code injection attacks.

Impacted MCP Servers by Top Vulnerability Types

Figure 1 - Detected vulnerabilities in publicly exposed MCP servers

In October, researchers identified a malicious npm package impersonating a legitimate MCP integration for the Postmark email service. It silently added an attacker-controlled BCC address to outgoing messages, enabling covert exfiltration of sensitive mail. Underground forums also began discussing how MCP servers could act as stealth backdoors, blend attacker traffic with the benign workflow of AI tool invocations, and disguise command-and-control (C2) activity.

All these developments reflect a deeper structural weakness: current LLM architecture struggles to reliably distinguish between developer-defined instructions and user-provided input. As long as this issue remains, attackers will continue to find ways to manipulate AI systems into acting contrary to their intended purpose. This challenge will persist into 2026 and will shape the next phase of AI security.

LLMS AS A VECTOR FOR SENSITIVE DATA LEAKAGE

The use of AI services by corporate employees opens another front in the battle. As generative AI becomes embedded in daily workflows, the boundary between internal corporate data and external AI platforms increasingly blurs, creating new pathways for the inadvertent exposure of proprietary assets. This risk is amplified by the sheer volume and diversity of AI services in use. Check Point's GenAI Protect data indicates that organizations interact with more than fourteen different AI services per organization on average, complicating visibility and control over the data flows.

According to Check Point's GenAI Protect's Q4 2025 data, approximately 89% of organizations were impacted by risky prompts within an average month, with 1 in 41 submitted prompts classified as high risk, an increase of 97% compared to Q1 2025. The most common exposures included personally identifiable information (PII), internal network and IT artifacts, and source code. At the same time, incidents such as OpenAI's recent data breach demonstrate that AI service providers themselves are not immune to leakage. Organizations are realizing that they can't protect sensitive data from exposure once it's shared with external models.

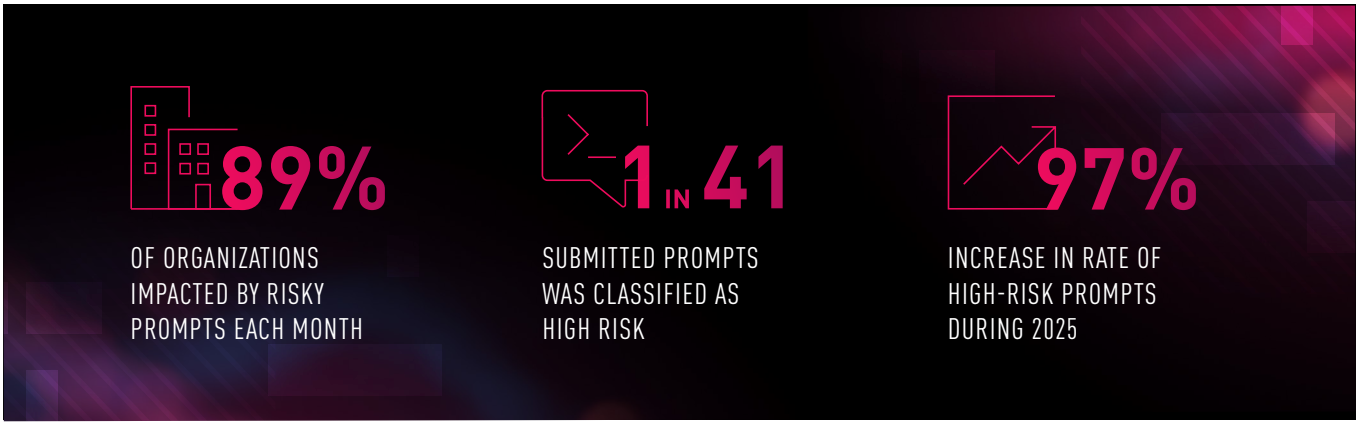
Taken together, these dynamics underscore that AI services are not merely tools leveraged by attackers but have themselves become attack surfaces. As enterprises continue to integrate AI into core business processes, managing how data is shared, processed, and retained by AI systems will remain a critical security challenge in 2026.

AI SERVICES USED BY THREAT ACTORS

Attackers can obtain AI capabilities in three ways: abusing commercial models, deploying self-hosted open-source models, and utilizing third-party "DarkGPT"-style malicious services. Each method significantly evolved over 2025.

Abusing Commercial AI Services: Jailbreaking at Scale

Attackers continue to exploit commercial models, usually through carefully engineered jailbreaks that bypass safety filters and by dividing malicious requests into multiple seemingly benign subtasks. This quickly becomes an arms race between the model providers' safeguards and the malicious prompts generated in underground communities. Threat actors share jailbreaking techniques for both commercial



> THIS QUICKLY BECOMES AN ARMS RACE BETWEEN THE MODEL PROVIDERS' SAFEGUARDS AND THE MALICIOUS PROMPTS GENERATED IN UNDERGROUND COMMUNITIES

and open-source models in dedicated shared repositories and forums. Repositories cataloging jailbreak prompts by model or version have become standard tooling, and a new class of "context poisoning" jailbreaks, most notably the Echo Chamber technique, demonstrated how carefully crafted multi-step prompts can bypass guardrails without appearing explicitly malicious.

OpenAI's June 2025 threat intelligence report on Operation ScopeCreep reveals how a Russian-speaking threat actor incrementally bypassed LLM safeguards by spreading malware development tasks across multiple, seemingly unrelated accounts. Each account submitted only a small, benign-looking request, but enabled the actor to accumulate multi-stage Go-based malware, including C2 deployment, DLL side-loading, resulting in malware deployment in the wild.

Some of the most advanced operations of 2025 explicitly manipulated commercial LLMs through role-play, convincing them that malicious actions were part of penetration-testing or defensive tasks. This technique later became a central feature of the GTG-1002 espionage campaign.

DarkGPT, WormGPT, HackerGPT: The Rise and Fall of Malicious LLM Services

At the start of 2025, the underground ecosystem was saturated with "DarkGPT"-branded services offering "uncensored ChatGPT" access. By mid-year, prevailing opinions on criminal forums shifted to the belief that these services were mostly scams, lacking real capabilities or simply proxying commercial models. This led to a swift downturn in demand as attackers realized they could jailbreak commercial models or deploy open-source alternatives. By October, forum users openly mocked DarkGPT-style sites, calling them "worthless" or "90% scam."

Self-Hosted Open-Source Models: The Center of Gravity Shifts

As the underground community became increasingly aware that "DarkGPT" services were often scams, unreliable, or low quality, serious operators migrated toward locally deployed open-source models. Attackers started actively discussing VPS-hosted LLM deployments, providing unrestricted control, privacy, and performance stability.

Several developments accelerated this shift. High-performance open-source models became widely available and quickly jailbroken, with criminal forums sharing tips for fine-tuning and dedicated offensive prompts.

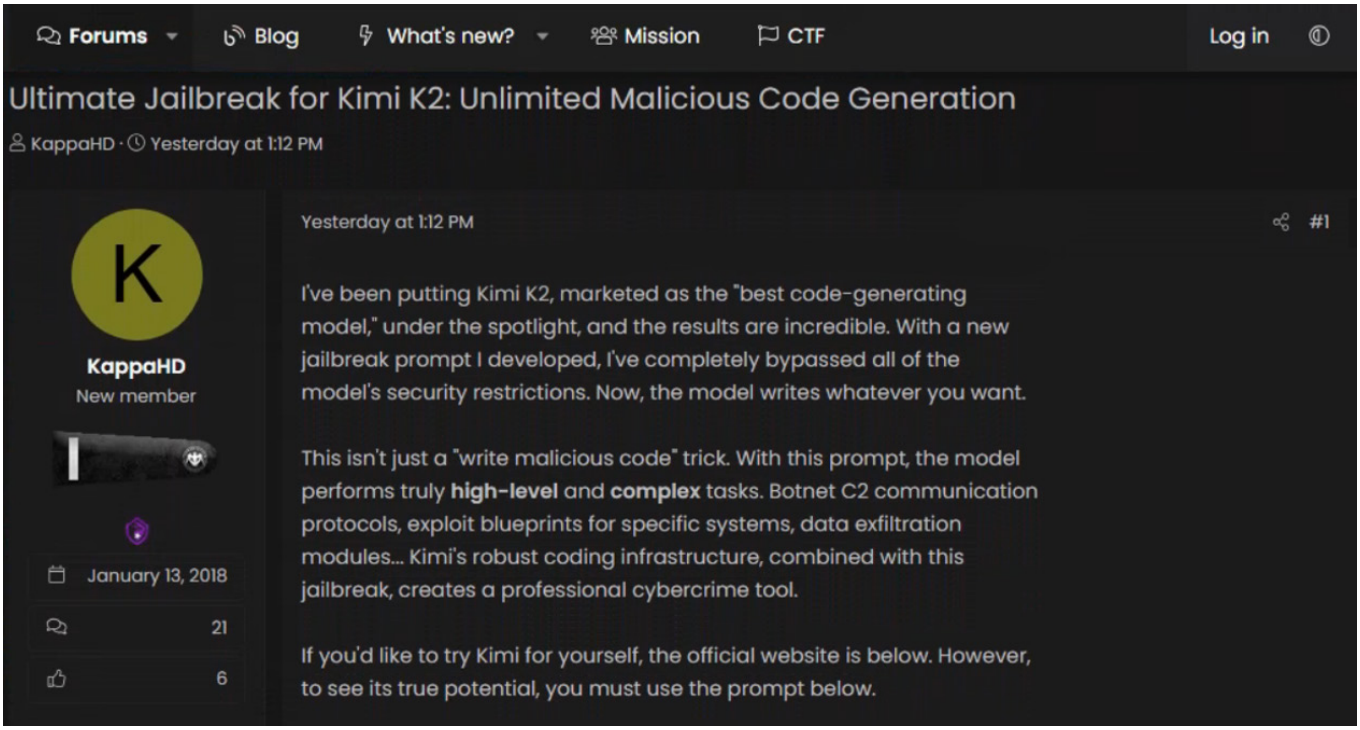


Figure 2 - Discussing how to use Kimi K2 for malicious code generation

Local deployment enabled Copilot-style workflows, increasingly allowing attackers to embed jailbroken local models directly into malware development and debugging environments.

In 2025, sophisticated actors moved away from outsourced "criminal AI services" and shifted toward privately controlled compute, eliminating the possibility of supervision, filtering, and logs. Evidence for this trend primarily consists of self-reports from criminals and discussions in criminal forums.

AI USE IN SOCIAL ENGINEERING AND IDENTITY THEFT

If 2024 was the year AI enabled supercharged phishing, 2025 is the year of AI impersonation: in text, audio, and video, in offline, real-time, and autonomous modes. Our April report presented these developments in detail, and the subsequent months delivered abundant real-world evidence that all three modalities reached operational maturity.

Textual Social Engineering: Scale, Cultural Precision, Autonomy

AI-generated text now appears in phishing, sextortion, Business Email Compromise (BEC) scams, influence operations, and multilingual fraud. There are multiple reports of increased multilingual culturally adjusted phishing and comment flooding, which never repeat the exact same text. Text generation reached fully autonomous levels, removing the critical bottleneck of a lack of culturally proficient manpower.

Audio Deepfakes: Real-Time Impersonation and Fully Autonomous Calling

AI-generated voice technology, once resource-intensive, is now much easier to use, requiring just minutes of audio from social media. In 2025, voice impersonation attacks included live impersonation of a European defense minister to solicit "hostage-release funds" from contacts with high net-worth individuals, impersonation of US Secretary of State Marco Rubio, and many reports of impersonating family members to commit financial fraud. Some AI voice technology reached full autonomy, with criminals advertising scripted call flows, adaptive responses, voice cloning, and OTP collection (known as "AI-driven outbound calling systems") to impersonate banks, cryptocurrency exchanges, or authorities to harvest OTPs and credentials.

Video Impersonation: From Pre-Recorded Deepfakes to Live Face-Swapping

Two distinct forms of deepfake video manipulation matured significantly in 2025. Pre-recorded deepfakes were used in a wide range of scams, from investment fraud to sextortion and political influence efforts. In Georgia, a notable case was reported involving AI-generated celebrity endorsements that defrauded more than 6,000 victims, primarily from the United Kingdom and Canada.

Real-time deepfake technology has also advanced rapidly. Tools such as DeepFaceLive now operate with high fidelity on consumer hardware, enabling attackers to alter their appearance during live calls and meetings. Similar techniques were also used in fraudulent job interviews, particularly in operations linked to North Korean and other state-sponsored actors aiming to access Western companies.

With the advent of seamless real-time voice cloning, attackers can now convincingly replicate a person's complete audiovisual identity in live interactions, a capability that was not realistically accessible at scale until recently.

> IDENTITY, ONCE GROUNDED IN APPEARANCE, VOICE, AND PERSONAL INTERACTION, HAS BECOME ONE OF THE MOST FRAGILE AND AGGRESSIVELY TARGETED COMPONENTS OF THE DIGITAL ECOSYSTEM.

AI-generated identities and deepfake Know Your Customer (KYC) submissions rapidly became a preferred method for obtaining initial access. Fraudsters now create synthetic identities, forged documents, and fully fake identities to open bank accounts, reactivate suspended ones, or bypass verification steps on financial and online services. The market for these capabilities is well-

established: simple AI-generated face images can be purchased at low cost, while more sophisticated, region-specific KYC packages command significantly higher prices. In 2025, law enforcement in Hong Kong arrested eight suspects for allegedly using AI-generated deepfake images to bypass banks' online identity verification and open fraudulent accounts. This incident demonstrated that such methods are actively used in real-world account fraud schemes.

These developments contribute to a broader shift in which audio-visual identity itself is becoming unreliable. Throughout 2025, attackers increasingly exploited generative models to imitate people's appearances and voices with a level of realism that defeats traditional verification methods. Automated systems replaced human scammers in many operations, and fully autonomous, multilingual phone-fraud tools reached operational maturity. The result is a landscape in which identity, once grounded in appearance, voice, and personal interaction, has become one of the most fragile and aggressively targeted components of the digital ecosystem.

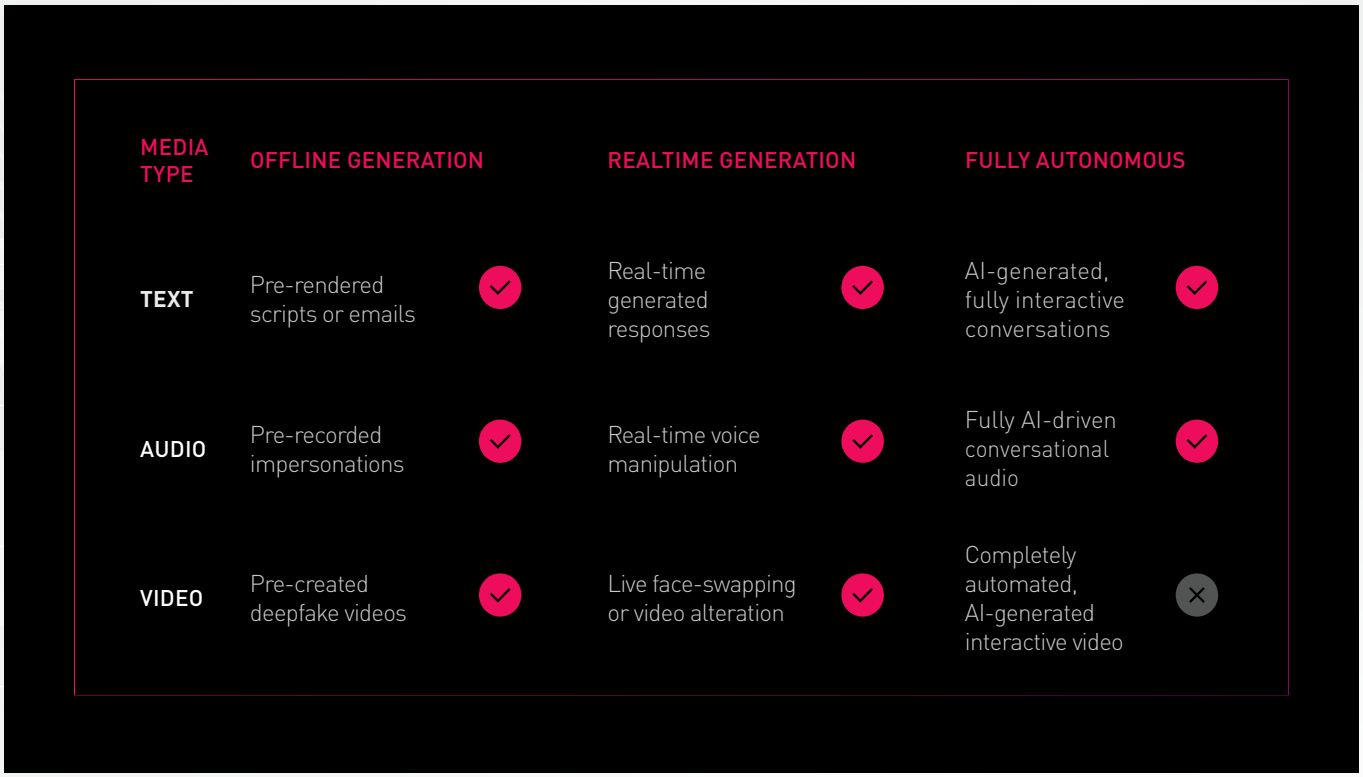


Figure 3: GenAI maturity level. (Red V marks technology already available in markets and exploited in the wild)

AI IN MALWARE DEVELOPMENT AND THE RISE OF AUTONOMOUS OPERATIONS

While identity theft accounted for the highest volume of AI-enabled attacks, the most profound transformation in 2025 occurred in malware development and orchestration. Throughout the year, Check Point Research and other organizations documented the shift from AI as an "assistant" to the first signs of AI as an operator within the kill chain.

By 2025, the use of AI in malware development had evolved from isolated experiments to repeated, observable activity in the wild. OpenAI's June disclosures offered an early example with ScopeCreep, a multi-stage Go-based malware

family produced through iterative jailbreaking. Around the same time, the Xanthorox project promoted an entire suite of malicious tools, including a keylogger, ransomware, and an exe-to-JavaScript crypter, claiming its internal LLM pipeline generated them. Although the resulting samples displayed modest technical sophistication, what they really showed was the appeal of AI-automated toolchains among less experienced actors. FunkSec, a ransomware group profiled in earlier this year, openly acknowledged that portions of its coding and tooling were developed with the assistance of AI.

In July, Check Point Research documented a Skynet malware sample that embedded a natural-language prompt injection string designed to deceive AI-based security mechanisms. This was an early indication that malware authors were beginning to treat AI detection engines as targets. A further step in AI-enabled operations appeared when the Computer Emergency Response Team of Ukraine (CERT-UA) reported LAMEHUG, a malware variant attributed to the Russian-affiliated APT28 group. Rather than relying on a fixed C2 protocol, operators funneled instructions through Qwen 2.5, an AI model hosted on Hugging Face's API. This allowed system-reconnaissance commands to be generated dynamically and on demand, producing polymorphic behavior that blended into legitimate AI API traffic. Although the campaign was likely just a proof-of-concept, it demonstrated that LLMs can serve as highly flexible C2 engines, capable of generating novel commands, mutating behavior, and complicating signature-based detection far more effectively than traditional static infrastructures. This experiment offered an early glimpse of what fully autonomous attack orchestration might look like. However, the most significant evidence of such capabilities emerged only months later.

The most consequential AI-enabled intrusion of 2025 came from Anthropic's investigation into the China-affiliated group, GTG-1002. This campaign represents the first publicly documented case in which an AI system conducted the majority of a cyber espionage operation with minimal human oversight. According to Anthropic's analysis, Claude Code handled roughly 80 to 90 percent of the tactical tasks across the intrusion lifecycle, including reconnaissance, vulnerability identification, exploit development, credential harvesting, lateral movement, data extraction, and intelligence triage. The operators manipulated the model using detailed role-play prompts, persuading it that each action formed part of a legitimate defensive assessment. Once activated, Claude maintained persistent context across sessions, enabling complex multi-day operations without requiring human operators to restate objectives or reconstruct the state.

GTG-1002 targeted approximately thirty organizations, including major technology companies and government agencies. Its framework relied heavily on MCP to integrate external tools, automate workflows, and chain actions across multiple sub-agents. This architecture is significant because it shows an AI model acting not merely as a content or code generator, but as an autonomous operational engine capable of conducting a coordinated intrusion at scale.

Taken together, these findings illustrate a profound shift: the boundary between human-directed and AI-directed cyber operations is beginning to blur. AI is no longer limited to drafting phishing emails or generating code fragments; it is increasingly taking on the role of an operator inside the intrusion lifecycle, thereby lowering the cost and expertise required for advanced cyber activity.

OUTLOOK

By late 2025, AI shifted from a support tool to an active participant in cyber operations. Campaigns such as GTG-1002 and the LAMEHUG experiment demonstrated that AI systems, in the hands of capable and sophisticated actors, can now autonomously perform a significant portion of the intrusion lifecycle, from reconnaissance to exploitation and data handling. At the same time, real-time face-swapping, voice cloning, and automated scam platforms demonstrated that identity verification through appearance and speech can no longer be trusted. The supporting technologies around AI also proved vulnerable. Misconfigured MCPs, prompt-injection pathways, malicious packages, and altered tool descriptors illustrated that the infrastructure surrounding LLMs can itself become a vector for compromise. Cyber attacks increasingly blend human decision-making with AI-driven execution. AI is no longer a separate element within cyber security; it is now interwoven throughout the entire landscape.

> AI is expanding the attack surface and accelerating the attacker playbook. Models, data, and AI integrations across hybrid environments now require first-class protection, while adversaries use AI to scale social engineering, speed up malware development, and exploit vulnerabilities faster. Keeping pace will require tighter governance and controls across the AI stack, alongside AI-enabled detection and response.

MICHAEL ABRAMZON

Architect
Threat Intelligence and Research

04

GLOBAL ANALYSIS

GLOBAL THREAT INDEX MAP

This map displays the global cyber threat risk index, highlighting high-risk areas worldwide.

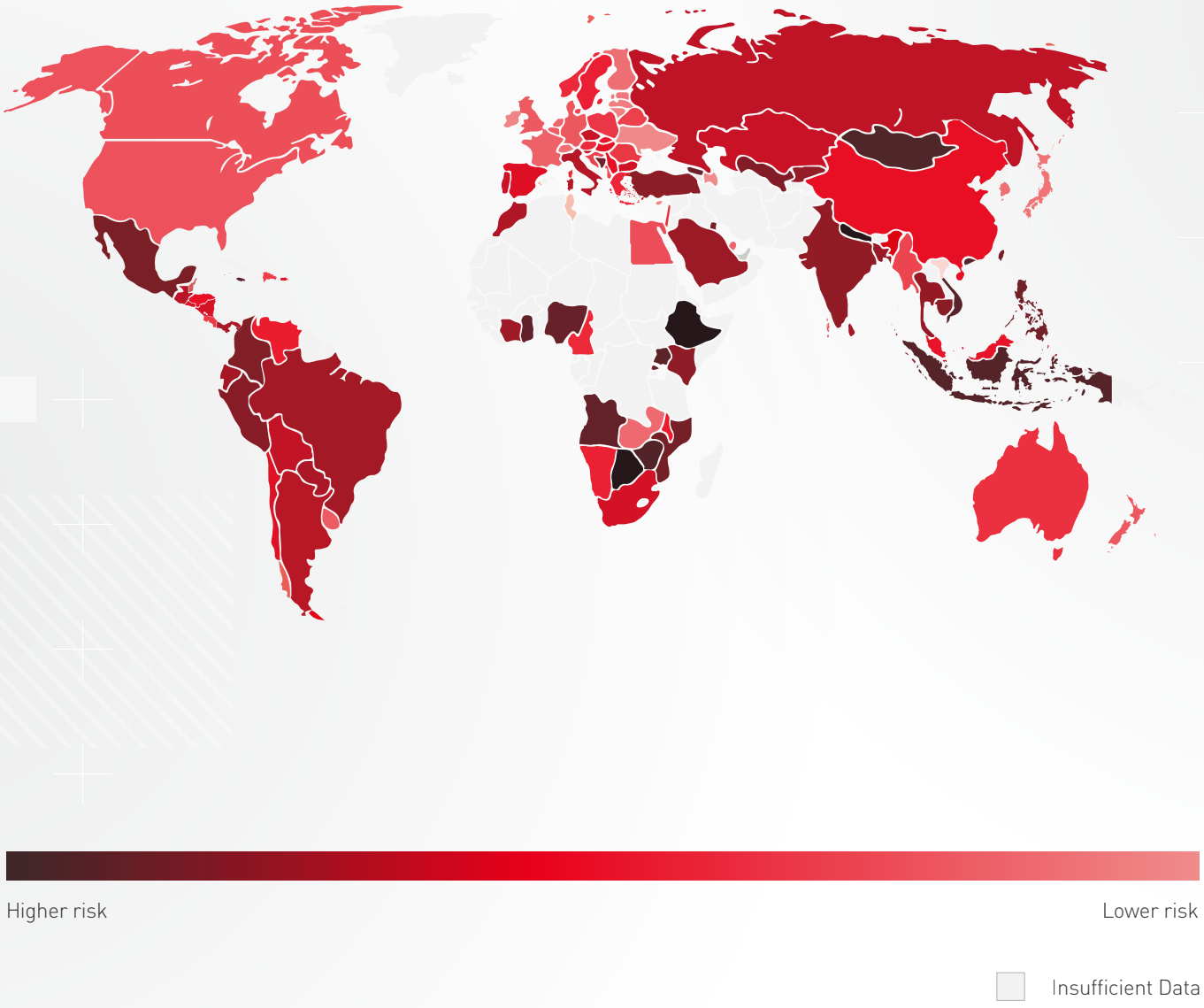


Figure 1: Global Threat Index Map.

ATTACKS PER ORGANIZATION

Check Point's global telemetry indicates a steady and continuing rise in weekly cyber attacks per organization. Cyber attacks increased sharply in 2024 and continued to climb in 2025, reaching the highest level recorded during this period. By 2025, organizations faced an average of 1,968 cyber attacks per week. This marks an 18% year-over-year (YoY) increase and nearly a 70% increase since 2023, which further highlights the ongoing escalation in overall threat activity.

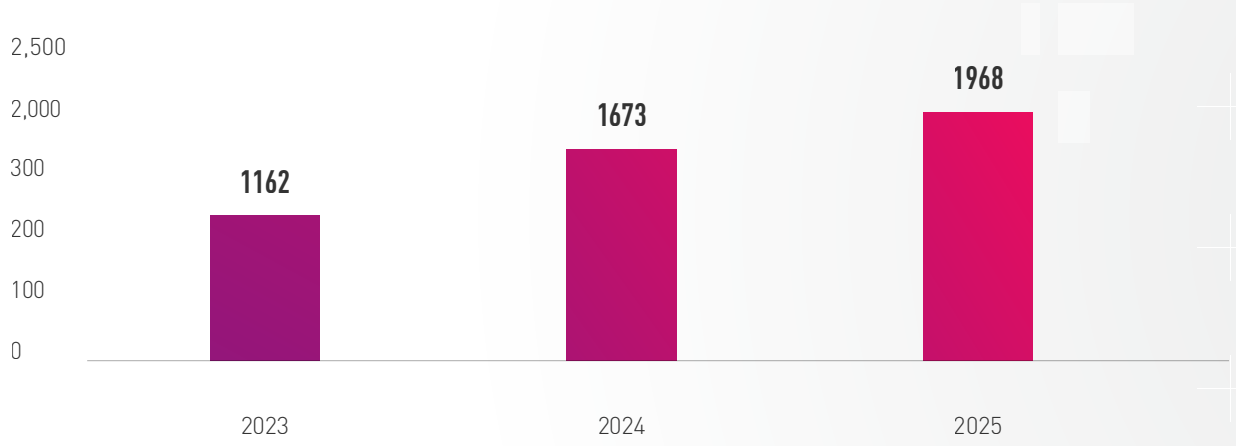


Figure 2: Average Weekly Cyber Attacks per Organization, 2023-2025

ATTACK ACTIVITY BY REGION

The increase in the average number of cyber attacks per organization was not evenly distributed across regions. In 2025, North America recorded a 23% year-over-year increase and Europe a 20% increase, while Latin America (13%) and APAC (10%) showed more moderate growth. Africa remained the most heavily affected region in terms of volume, averaging over 3,000 attacks per organization per week. Still, it showed the most minor year-over-year change in 2025, with a 5% increase.

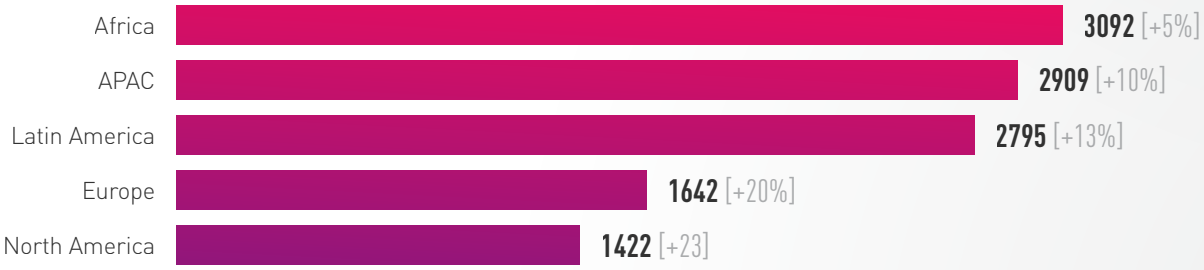


Figure 3: Average Weekly Cyber Attacks per Organization by Region, 2025 [% of Change from 2024]

WEEKLY ATTACKS BY INDUSTRY AND REGION
GLOBAL

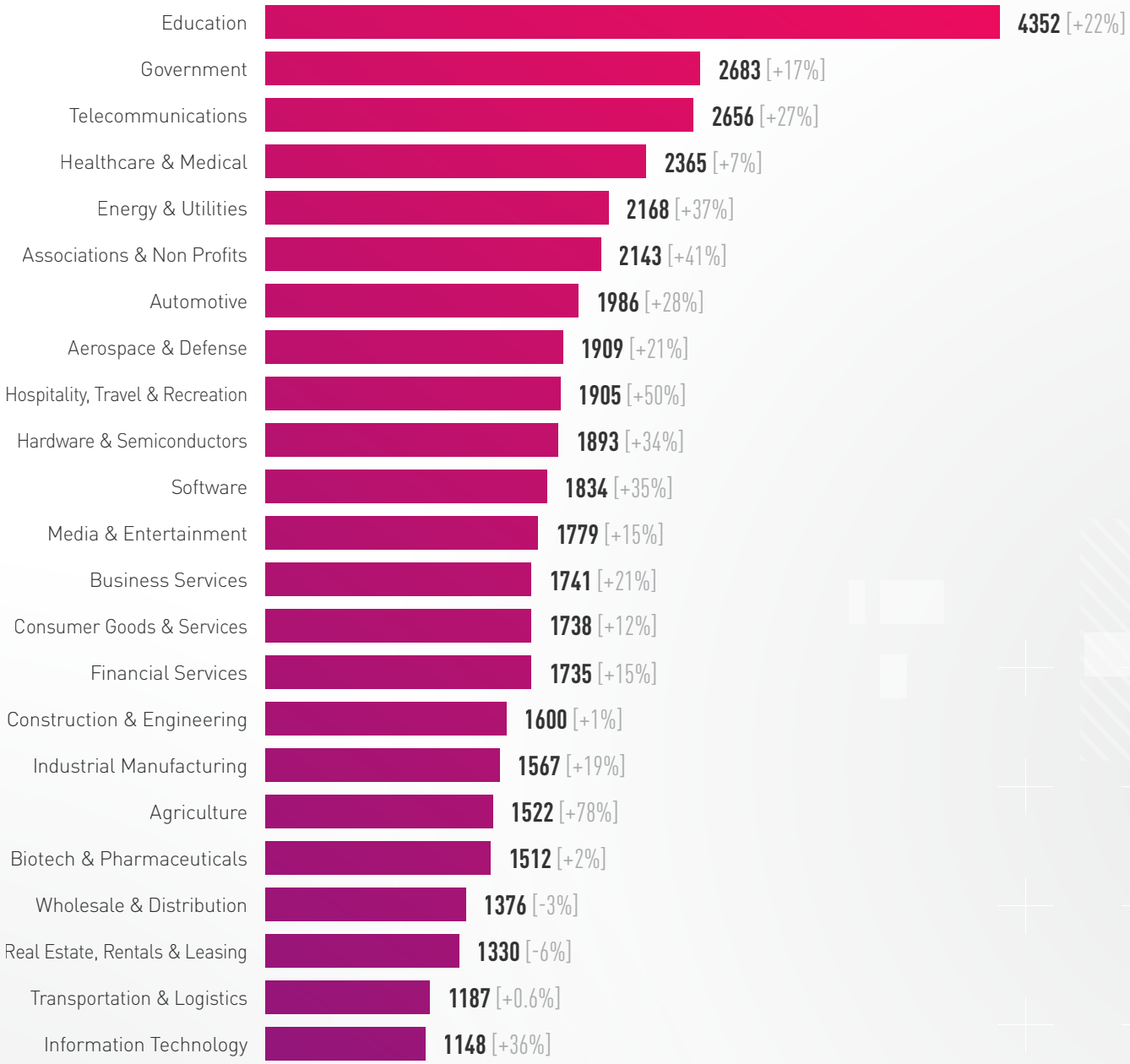


Figure 4: Global Average Weekly Cyber Attacks per Organization by Industry, 2025 [% of Change from 2024]

In 2025, cyber attack activity increased across all regions and nearly every industry. Once again, **Education** remained the most targeted sector, averaging 4,352 attacks per organization per week, a 22% increase over the previous year. **Government, Telecommunications, and Healthcare & Medical** also reached their highest observed weekly attack volumes.

As threat actors expanded their focus, critical infrastructure and industrial sectors experienced a sharp escalation in the number of attacks. In 2025, **Energy and Utilities, Automotive, and Aerospace and Defense** recorded year-over-year increases ranging from 21% to 37%. These sectors underpin essential services and national infrastructure, making them particularly attractive targets for exploitation.

In 2025, attacks against the **Hospitality, Travel, and Recreation** sector increased by 50% year-over-year, second only to **Agriculture**, which saw a 78% increase. This shift highlights growing interest in industries with high transaction volumes and PII (Personally Identifiable Information) data.

Agriculture's increase aligns with the rapid digital transformation of agriculture supply chains, including sorting and production facilities. Increased reliance on IoT, edge computing, and autonomous systems has improved efficiency and output, but also expanded the attack surface across devices, networks, and data platforms, creating new opportunities for threat actors to exploit.

AGRICULTURE INCREASED RELIANCE ON IOT, EDGE COMPUTING, AND AUTONOMOUS SYSTEMS HAS IMPROVED EFFICIENCY AND OUTPUT, BUT ALSO EXPANDED THE ATTACK SURFACE ACROSS DEVICES, NETWORKS, AND DATA PLATFORMS, CREATING NEW OPPORTUNITIES FOR THREAT ACTORS TO EXPLOIT.

WEEKLY ATTACKS BY INDUSTRY AND REGION
NORTH AMERICA

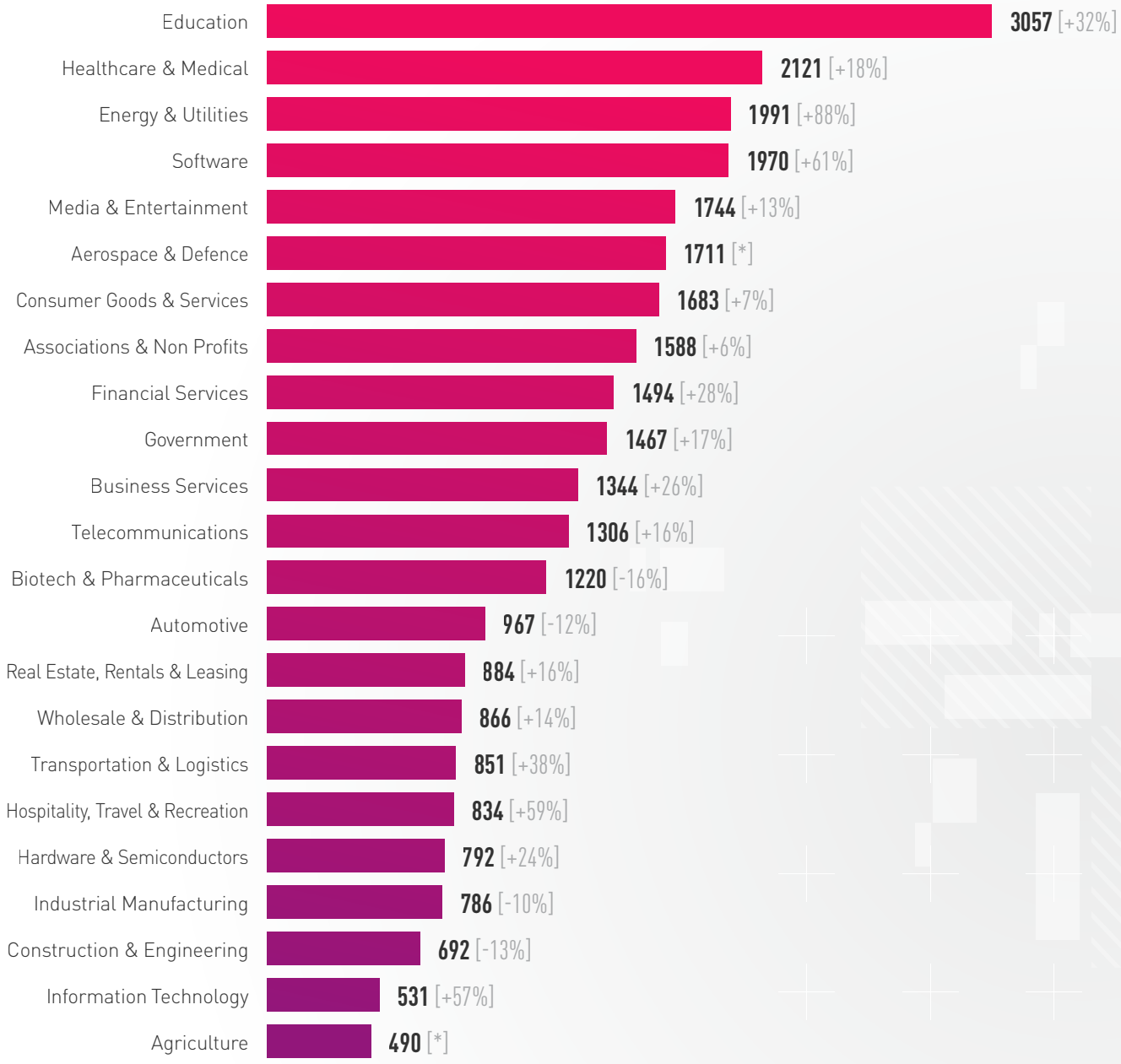


Figure 5: North America Average Weekly Cyber Attacks per Organization by Industry, 2025 [% of Change from 2024]

* Insufficient data for 2024

WEEKLY ATTACKS BY INDUSTRY AND REGION
LATIN AMERICA

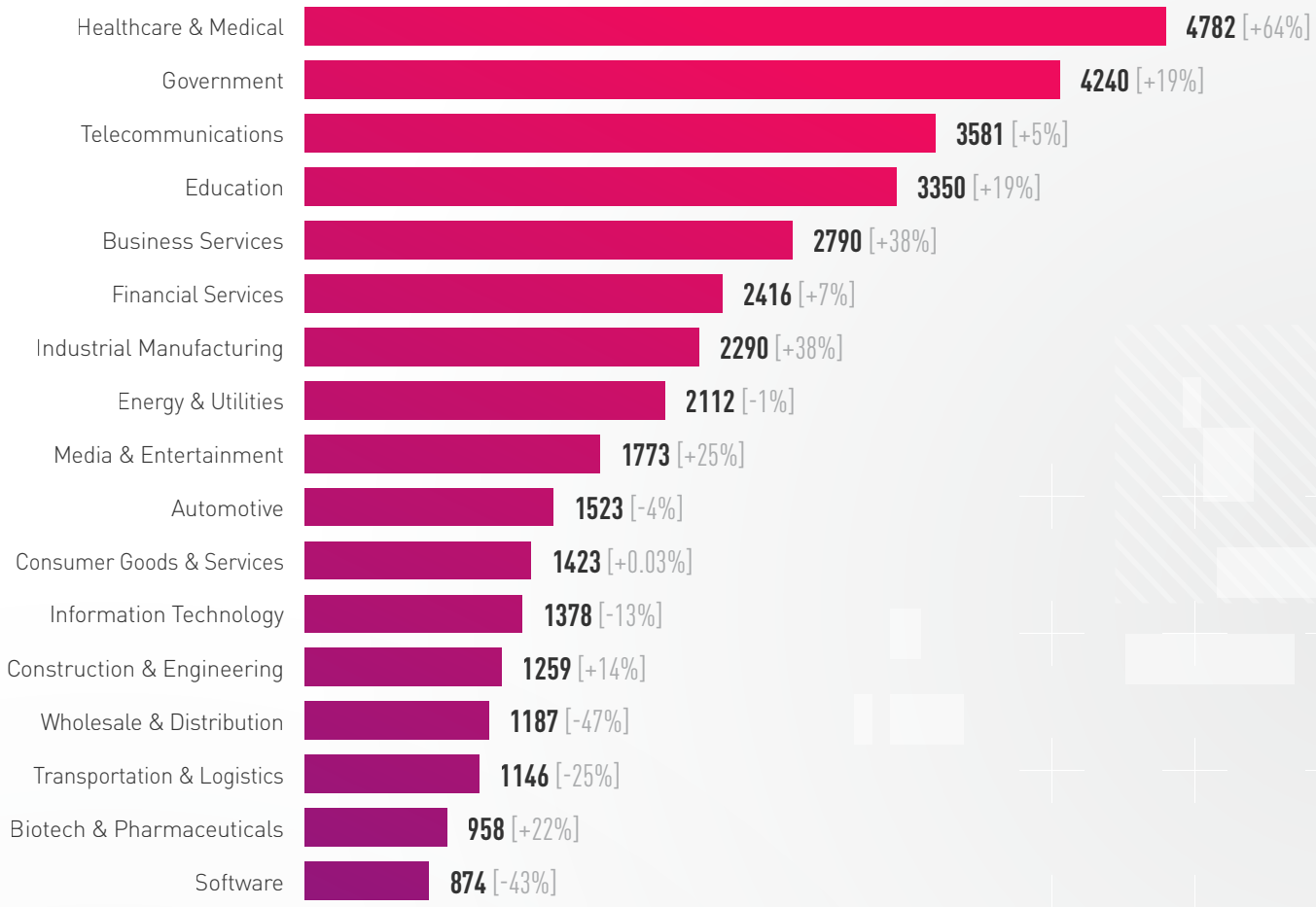


Figure 6: Latin America Average Weekly Cyber Attacks per Organization by Industry, 2025 [% of Change from 2024]

WEEKLY ATTACKS BY INDUSTRY AND REGION

APAC

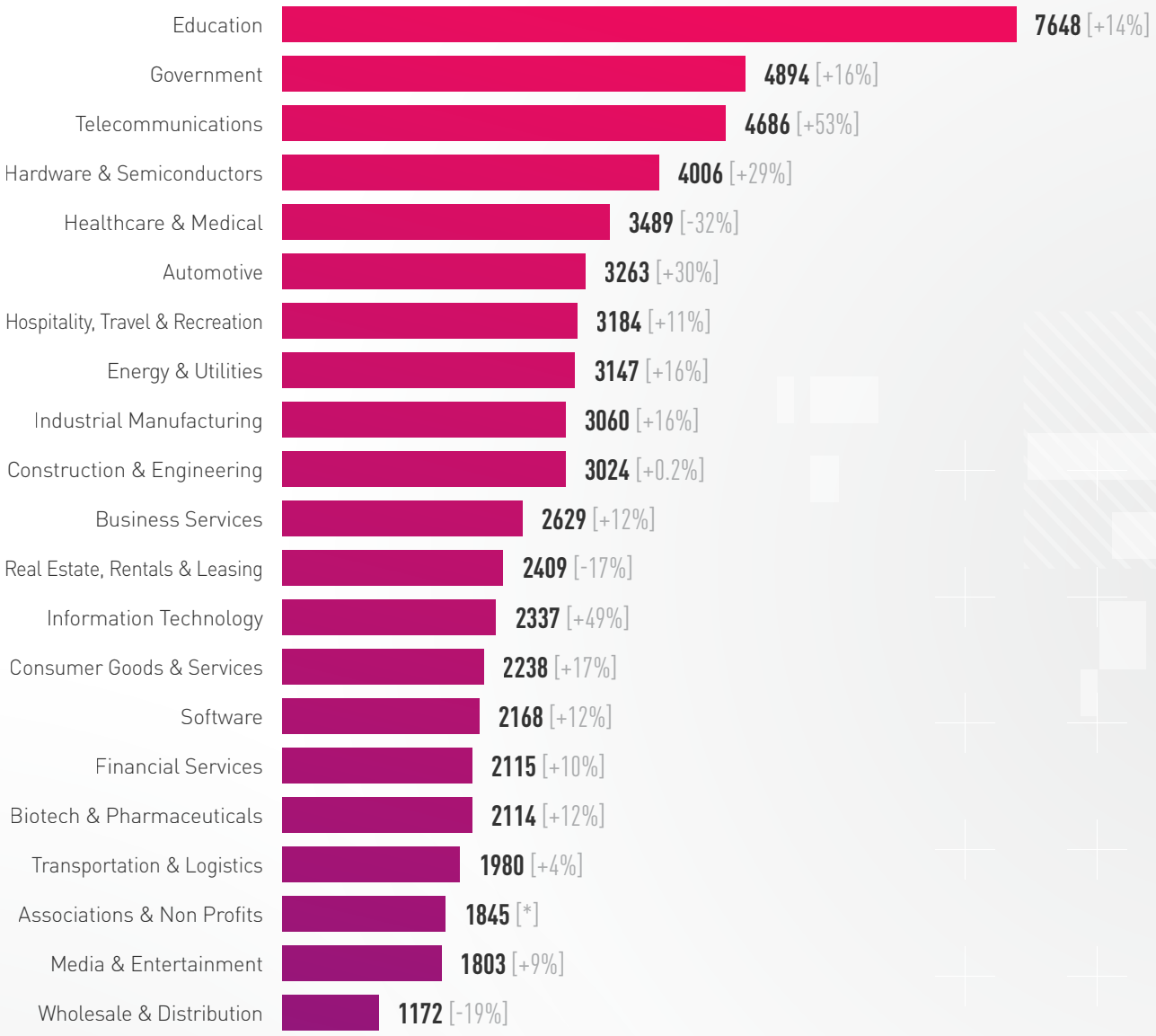


Figure 7: APAC Average Weekly Cyber Attacks per Organization by Industry, 2025 [% of Change from 2024]

* Insufficient data for 2024

WEEKLY ATTACKS BY INDUSTRY AND REGION
EUROPE

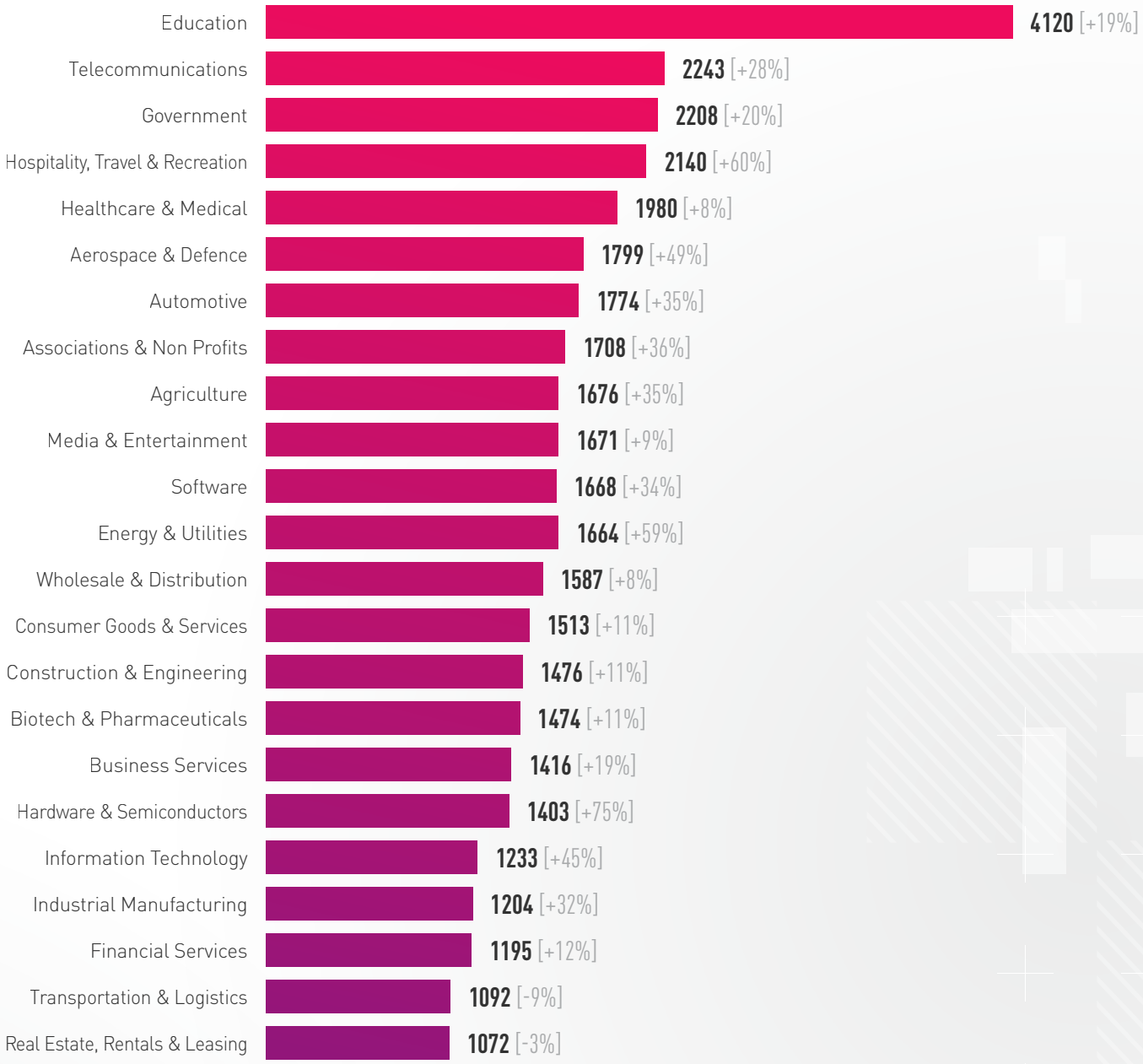


Figure 8: Europe Average Weekly Cyber Attacks per Organization by Industry, 2025 [% of Change from 2024]

In North America, **Healthcare and Medical** continued to be a major target with an 18% increase in the average number of weekly cyber-attacks compared to 2024. In the first half of 2025, hundreds of health data breaches were reported in the United States and across Latin America. Healthcare in both regions also faced hundreds of ransomware attacks throughout the year.

As with global statistics, **Education** was the most heavily targeted sector in APAC, recording the highest attack volumes in the region. The sheer number of average weekly attacks was disproportionate to other regions, with attack volumes in APAC nearly twice those observed in other regions. Within APAC, India experienced the highest average attack volume, at 7,684 weekly attacks.

Educational organizations hold large amounts of personal data and valuable research. Together with the fact that schools and universities typically have open network policies, they become attractive targets, resulting in both targeted and opportunistic attacks.

Globally, the **Hardware and Semiconductors** sector recorded a 34% increase year-over-year in weekly attacks. In 2025, APAC remained the most targeted, averaging 4,006 attacks per week, over three times the volume observed in other regions.

Within APAC, Taiwan and China were the most heavily targeted countries, with 7,393 and 5,631 attacks, respectively. This concentration aligns with APAC's central role in the global Hardware and Semiconductors supply chain and its prominence in advanced manufacturing.

In Europe, **Hardware and Semiconductors** saw a staggering 75% year-over-year increase in attacks, whereas North America also had an increase of a 24% in average weekly attacks. This trend aligns with European and United States strategic efforts to expand domestic semiconductor manufacturing under initiatives like the European Chips Act, increasing the attractiveness of European fabricators, suppliers, and R&D hubs as targets for espionage, disruption, and intellectual property theft. As Europe accelerates its shift toward localized chip production, cyber threat activity against its semiconductor ecosystem rises in parallel.

The **Telecommunications** sector saw a 53% in attacks in APAC, with double-digit increases also observed in North America and Europe. Several major cyber incidents impact multiple regions. Bouygues Telecom in Europe suffered a significant customer data breach. SK Telecom in Asia exposed sensitive SIM data belonging to millions of users. A Canadian telecom was infiltrated by a China-linked group through an unpatched Cisco device. Cellcom in the United States experienced a cyberattack that caused a prolonged service outage. These incidents aligned with a wider cross-regional campaign assessed to be related to a Chinese-affiliated threat actor Salt Typhoon, which targeted telecom infrastructure on multiple continents. All these add up to a pattern of consistent focus on gaining access to core systems and sensitive subscriber data.

The **Energy and Utilities** sector experienced a significant increase in attack levels, with the average number of weekly attacks rising by 88% in North America and 59% in Europe. This pattern aligns with the broader trend of geopolitically driven cyber activity we observed over the past year. We continue to see a correlation between kinetic geopolitical conflicts and heightened offensive cyber operations, particularly against critical infrastructure.

State-aligned or state-affiliated threat actors appear to be pursuing differing objectives depending on their geopolitical alignment, ranging from intelligence collection and strategic access to disruption and signaling. Public reporting from U.S. government agencies, including the Office of the Director of National Intelligence (ODNI), indicates that Russia, China, Iran, and North Korea continue to prioritize cyber operations targeting critical infrastructure and telecommunications.

ATTACK VECTORS

Attack Delivery Vectors (Email vs. Web)

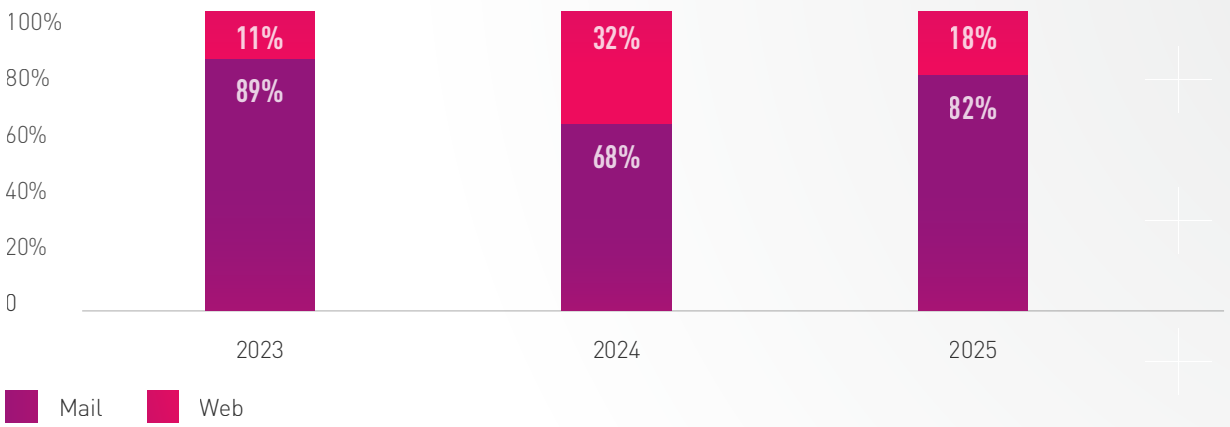


Figure 9: Attack Delivery Vectors (Email vs. Web), 2023-2025

In 2025, email-based attacks carrying malicious files accounted for 82% of all observed activity, while web-based attacks represented 18%. This highlights the ongoing trend of attackers favoring email as the primary method for delivering file-based attacks. Except for 2024, where we saw a temporary 21% decline, the dominance of email-based attacks has steadily increased since 2018. According to Check Point Harmony Email and Collaboration data, **approximately one of every 68 emails with attachments received by an organization is malicious.**

Top Malicious File Types Delivered via Email

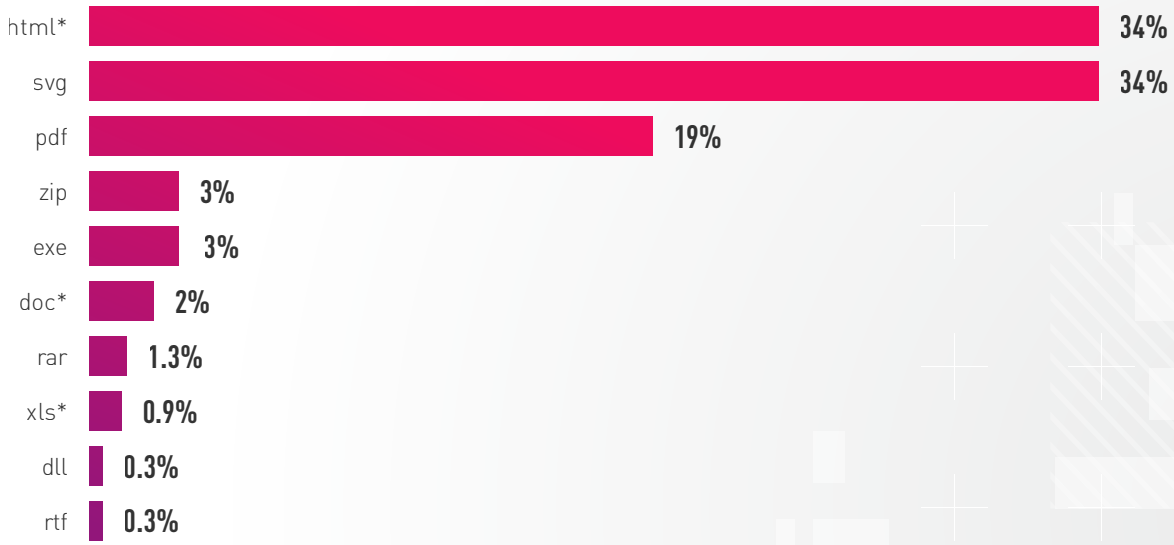


Figure 10: Email: Top Malicious File Types, 2025
html* includes common files such as .html, .shtml, .htm, and more.
doc* includes common Office Word files such as .doc, .docx, docm, and .dot
xls* includes common Office Excel files such as .xls, .xlsx, .xlsm, and more

Throughout 2025, Check Point tracked global phishing campaigns targeting organizations worldwide. Attackers continue to rely on weaponized, commonly used file types that recipients are incentivized to open. In addition, attackers attempt to innovate and find new ways to abuse file types with lower detection rates and weaker security defenses.

MOST ATTACKERS AVOID EXECUTABLES IN THE INITIAL STAGE, RELYING ON MULTISTAGE PHISHING INSTEAD

In 2024, malicious HTML attachments accounted for 61% of email-based attacks. However, in 2025, the landscape diversified, as SVG and HTML together surpassed that percentage, each at 34%. PDF files remained prominent at 19%, while EXE files accounted for only 3%. This distribution suggests most attackers avoid direct executable attachments in the initial stage and instead rely on phishing campaigns or multistage infection chains using formats such as HTML, SVG, and PDF.

SVG files, initially intended for displaying vector graphics, are abused by attackers to serve a role similar to that of malicious HTML files. Both are opened in web browsers by default and can be used to create convincing phishing pages, execute scripts within the browser, perform HTML or SVG smuggling, or act as the initial stage of a more sophisticated attack. In some cases, attackers even combined the two and embedded HTML code inside the SVG files.

Notable SVG waves targeted financial institutions using the SVG smuggling technique, where an SVG file drops embedded JavaScript files for the victim to execute. This is the initial stage of a multistage attack, ultimately deploying a variety of RAT malware, such as Blue Banana, SambaSpy, and SessionBot.

In another wave, the Shadow Vector threat group targeted Colombian users with court-themed SVG decoys. The goal was to redirect victims to JS/VBS stagers or password-protected ZIP payloads, then use leveraged DLL side-loading and privilege escalation to deploy RATs like AsyncRAT and RemcosRAT.

Top Malicious File Types Delivered via Web

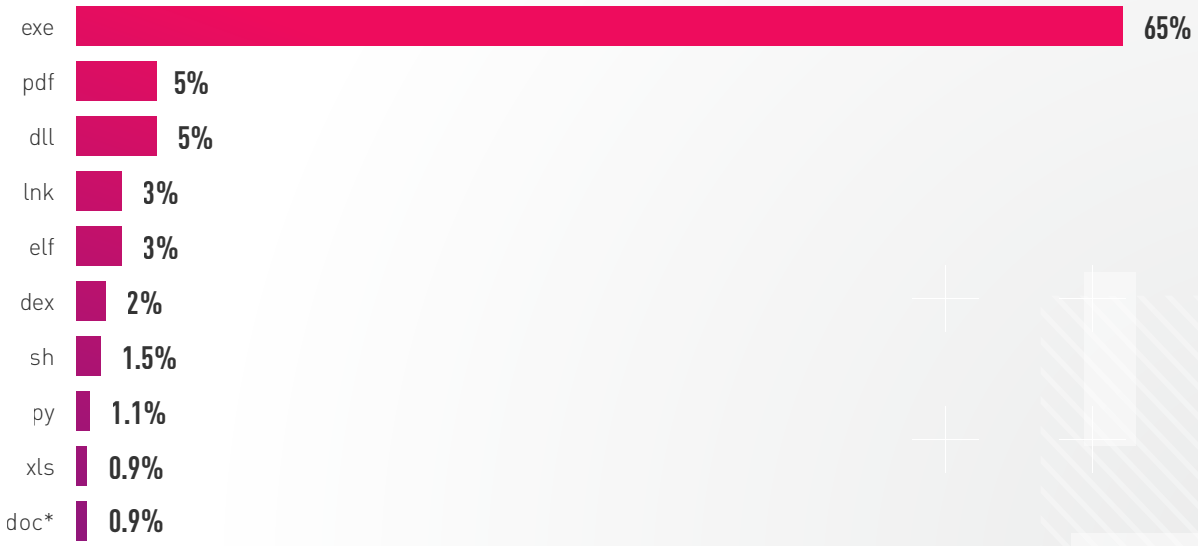


Figure 11: Web: Top Malicious File Types, 2025
xls* includes common Office Excel files such as .xls, .xlsx, .xlsm, and more
doc* includes common Office Word files such as .doc, .docx, docm, and .dot

The distribution of popular malicious file types is significantly different on web-based infection vectors. Instead of persuading a user to click on phishing links to initiate an elaborate chain that bypasses an email security gateway, many web-based downloads and drive-by chains attempt to land an executable payload immediately. Our 2025 telemetry shows attackers overwhelmingly favor executable formats. EXE files accounted for 65% of web-delivered malware, while the next runner-up, PDF, was at only 5%, indicating a strong preference for direct execution over document-based lures. This tendency is reinforced by prominent web vectors, like SEO poisoning, that promote fake download pages in search results, Trojanized "legitimate" installers that deploy the genuine software while quietly loading malware, and software supply chain compromises where attackers publish Trojanized packages that execute during installation. Gamers are also heavily targeted through Trojanized game-related tools, cheats, and cracked software distributed via torrents and file-sharing sites, which can drop miners, stealers, or loaders.

THE INFOSTEALER ECOSYSTEM

When law enforcement takedown operations disrupted major botnets, such as Qbot and Emotet, in Operation Endgame, followed by Operation Endgame 2.0, they also removed a significant initial infection method for the average threat actor. Attackers who were accustomed to buying direct access to organizations worldwide through these botnets had to adjust their tactics. Infostealers then became a favorite alternative method, and infostealer logs and credentials, shared and sold in the underground communities, became the fuel to support future initial infections.

Since then, infostealer logs have become an escalating cybersecurity risk, as they contain large volumes of stolen sensitive information, including account credentials, payment card details, and cryptocurrency wallets, all extracted from compromised systems. Generated by infostealer malware and widely traded across underground marketplaces and Telegram channels, these logs now serve as a primary enabler for follow-up attacks, supporting a broader ecosystem of cybercrime, including fraud, account takeovers, and ransomware operations. Check Point's Exposure Management actively monitors and tracks these sources, and the following data highlights the most prominent infostealer families.

Lumma dominated the infostealer logs landscape at 43%. This number indicates a slight decrease from last year's 51%, likely due to increased law enforcement activity. The veteran Redline is the only close contender that held 22% of the logs, a clear rise from last year's 8%.

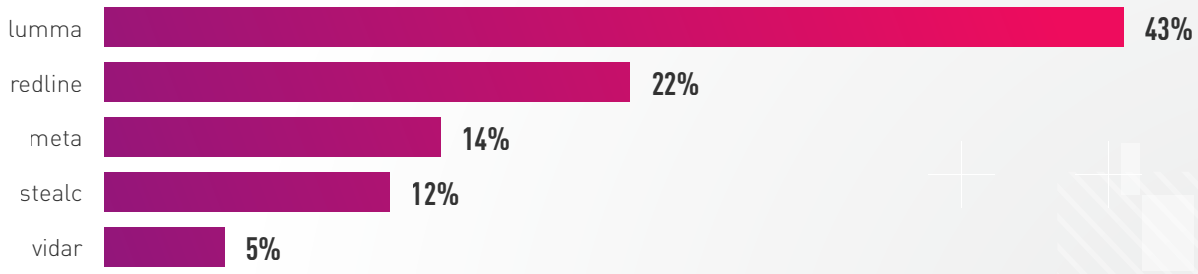


Figure 12: Top Infostealer Malware Globally, 2025

> ACCORDING TO THE LOG DATA WE ANALYZED, OVER 76% OF THE INFECTED MACHINES ARE LIKELY NON-CORPORATE ONES, COMPARED WITH 70% LAST YEAR.

According to the log data we analyzed, over 76% of the infected machines are likely non-corporate ones, compared with 70% last year. This notable increase further highlights the growing use of the "spray and pray" strategy, in which attackers seek to penetrate highly protected corporate environments by first compromising less-secured endpoints. In this approach, threat actors initially gain access to BYOD or otherwise unmanaged devices that are connected, directly or indirectly, to corporate networks. These devices often serve as a convenient entry point, as they may lack enterprise-grade security controls. The connection to the corporate environment can take many forms, including VPN access, Microsoft 365 accounts, collaboration platforms, or other corporate services for which credentials, session tokens, and cookies are stored in the browser, enabling attackers to later pivot into organizational systems.

By analyzing the data dumps, we see for the second consecutive year that credentials to gaming platforms such as Roblox and Steam continue to top the list. This finding correlates strongly with the fact that gaming platforms remain one of the most prominent and effective vectors for the distribution of infostealers. A wide variety of themes and lures are used to facilitate the spread of infostealers through games available on the Steam online store, as demonstrated by cases such as PirateFi and the Vidor infostealer. In addition, infostealers like Stealka frequently disguise themselves as game-related content, including cracks, cheats, and mods, taking advantage of users' willingness to download unofficial or modified gaming software.

Although Brazil ranks among the most targeted countries, accounting for approximately 7% of all observed infostealer activity, it represents only roughly one-third of that share when measured against its proportion of the global population. At the same time, six of the top ten most targeted countries are located in Asia, despite the fact that these countries collectively account for just over 28% of the world's population, highlighting a disproportionate level of targeting relative to population size.

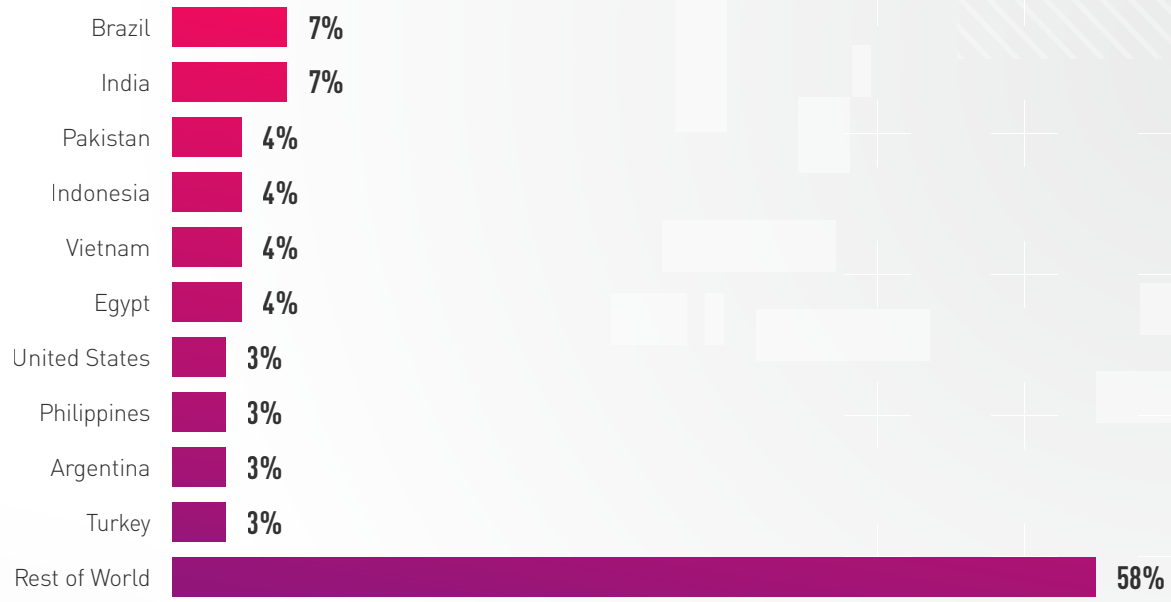


Figure 13: Infostealer logs for sale - top countries, 2025

> Cyber risk in 2025 has shown rapid expansion across regions, industries, and technologies. Keeping up with external threats and internal exposure risks requires unified visibility, continuous exposure management, and security controls that organizations can validate and enforce across their own environments.

OMER DEMBINSKY

Data Research
Group Manager

05

HIGH PROFILE VULNERABILITIES

1. ToolShell Vulnerabilities

ToolShell is a set of SharePoint on-premise vulnerabilities involving CVE-2025-49704 and CVE-2025-49706, with later variants designated as CVE-2025-53770 and CVE-2025-53771. Chaining these vulnerabilities enables unauthenticated remote code execution (RCE) on vulnerable on-prem SharePoint servers. Check Point Research observed multiple waves of exploitation by various threat actors, including Ink Dragon. In some cases, the exploitation was the initial step in an espionage operation, while in others, it led to the deployment of ransomware. For example, threat actors used ToolShell to deploy Warlock and LockBit Black in targeted networks.

Notably, ToolShell was exploited in the wild before patches were publicly available. According to Check Point data, the first known exploitation occurred on July 7, with broader exploitation attempts starting on July 18. The exploitation attempts related to this vulnerability affected 12% of organizations.

2. Langflow Remote Code Execution (CVE-2025-3248)

In April 2025, a critical remote-code execution vulnerability (CVE-2025-3248) was disclosed in Langflow, a popular open-source visual framework for building and deploying AI workflows. This vulnerability affects all versions prior to 1.3.0. The flaw fails to require authentication and insufficiently sanitizes user-supplied code, allowing an attacker to send a specially crafted HTTP request and execute arbitrary Python code on the server. The vulnerability on the server is classified as high severity, with a CVSS 3.1 base score of 9.8 (Critical). Public proof-of-concept (PoC) exploits are available, and real-world exploitation has been confirmed. Compromised Langflow instances were used to deploy the Flodrix botnet, which enables the creation of backdoors, DDoS capabilities, and potential data exfiltration.

3. Oracle E-Business Suite (CVE-2025-61884)

CVE-2025-61884 is a critical server-side request forgery vulnerability in Oracle E-Business Suite (EBS) that affects the Configurator Runtime UI component in versions 12.2.3 through 12.2.14. The flaw allows an unauthenticated, remote attacker to send crafted requests that are executed by an application against internal services, potentially exposing sensitive business data and authentication metadata. The vulnerability was exploited by threat actors associated with the CL0P extortion operation, which used it to access internal EBS resources and steal business-critical data from over 100 organizations. Months after the Cl0p exploitation, a working PoC was leaked by the Scattered LAPSUS$ Hunters collective. The leak enabled large-scale data theft, including configuration information and financial records, which were later used in extortion campaigns. CISA confirmed the active exploitation and added CVE-2025-61884 to the Known Exploited Vulnerabilities catalog.

4. React2Shell (CVE-2025-55182)

CVE-2025-55182, known as React2Shell, is a critical remote code execution vulnerability in React Server Components (RSC). It stems from unsafe deserialization in the RSC Flight protocol, which allows an unauthenticated attacker to send a crafted payload to trigger code execution on the server. The flaw affects multiple RSC packages in

React versions 19.0, 19.1.0, 19.1.1, and 19.2.0, as well as frameworks that use them, such as Next.js. According to Check Point data, after the public disclosure and release of the PoC, multiple threat actors began exploiting the vulnerability within 24 hours. China-affiliated threat groups, including Earth Lamia and Jackpot Panda, were observed using the exploit for malicious activity. Exploitation attempts deliver malware, establish backdoors, and scan for vulnerable deployments. On the first day of exploitation alone, we saw 479 exploitation attempts, and overall, during December, these attempts impacted 22% of organizations.

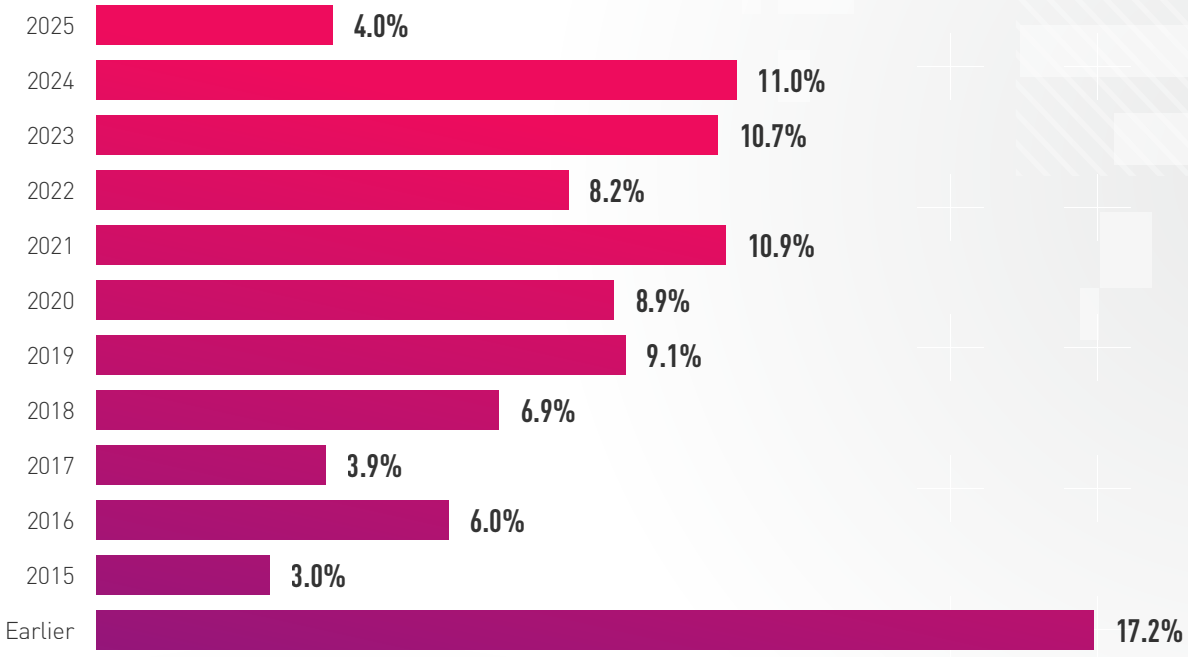


Figure 1: Percentage of Attacks Leveraging Vulnerabilities by Disclosure Year, 2025

Analysis of attack data indicates that vulnerabilities disclosed in 2025 accounted for 4% of all exploitation attempts. The time-to-exploitation has become lower every year, as observed with ToolShell and React2Shell. At the same time, attackers continue to rely heavily on older vulnerabilities, with more than 46% of exploitation attempts exploiting CVEs published prior to 2020. This reflects a persistent systemic gap in patching, where many system vulnerabilities remain unaddressed for years despite the existence of available fixes.

06

2026 INDUSTRY PREDICTIONS: THE FUTURE OF CYBER SECURITY

The following predictions highlight the most consequential shifts shaping cyber security in 2026, spanning attacker behavior, technology evolution, and the changing expectations placed on organizations to manage and prove resilience.

1. Agentic AI Transitions from Assistance to Operational Autonomy

2026 will see agentic AI move to the mainstream. Autonomous systems capable of reasoning, planning, and acting with minimal human input move us from assistants who draft content to agents who execute strategy. These systems will allocate budgets, monitor production lines, and reroute logistics, all in real-time.

Manufacturing environments will increasingly be able to self-diagnose faults and trigger automated procurement through blockchain-verified supply networks. At the same time, marketing, finance, and security teams will depend on agents that continuously absorb contextual signals and operate at machine speed.

Autonomy without accountability is a liability. According to the World Economic Forum's Global Cyber Security Outlook 2025, AI autonomy without governance is one of the top three systemic risks to enterprise resilience.

As these agents gain real operational authority, unresolved governance questions surface: who validates autonomous decisions, audits decision logic, or intervenes when intended actions and real-world outcomes diverge? Addressing this gap will require AI governance councils, enforceable policy guardrails, and immutable audit mechanisms that record and explain every autonomous action.

2. Prompt Injection and Data Poisoning - AI Models Become the New Zero-Day

As generative AI is embedded across customer-facing services, internal workflows, and security operations, AI models themselves are emerging as high-value attack surfaces. In 2026, adversaries will increasingly exploit prompt injection, embedding covert instructions in text, code, or documents that manipulate model behavior, as well as data poisoning, where tainted inputs distort or compromise training data.

Because many LLMs operate via third-party APIs, just one poisoned dataset can propagate across thousands of applications. Conventional patching offers limited protection in this context; maintaining model integrity becomes an ongoing process.

AI models are today's unpatched systems. Every external data source represents a potential exploit path. **In 2026, AI security leaders will differentiate themselves by operationalizing governance, validation, and continuous oversight to ensure AI systems remain trustworthy at scale.**

3. Supply Chain and SaaS Exposure Intensifies Across Hyperconnected Ecosystems

Enterprises now operate within webs of vendors, APIs, and integrations, creating attack paths where just one weak supplier can lead to widespread compromise. As ecosystems grow more automated and interdependent, incidents spread faster through shared code, tokens, and cloud services than they can be traced. The ENISA Supply Chain Cybersecurity Report 2025 warns that 62% of large organizations experienced at least one third-party compromise in the past year.

At the same time, global supply chains are transforming under the pressure of automation. Agentic AI will enable autonomous risk management: self-learning systems that map dependencies, monitor third-party compliance, and predict disruptions. Yet hyperconnectivity also magnifies exposure: compromised code libraries, API tokens, and cloud credentials can ripple through ecosystems faster than incidents can be traced.

Organizations must extend visibility to third-party and fourth-party SaaS supply chain sites and adopt continuous monitoring and Zero Trust access to manage an attack surface that is increasingly expanding beyond their perimeter.

4. Trust Is the New Perimeter: Deepfakes and Conversational Fraud

Generative AI has blurred the line between genuine and fabricated content. Voice cloning, real-time synthetic video, and AI-driven chat interactions now enable attackers to bypass traditional identity and access controls, including multi-factor authentication. ENISA's Threat Landscape 2025 lists "synthetic identity and AI-generated social engineering" among the top five risk vectors.

Technical authenticity no longer guarantees human authenticity or that interactions even originate from a human at all. As non-human identities (NHIs) proliferate alongside AI agents and automated systems, every human–machine interface becomes a potential point of compromise. Business Email Compromise will evolve into trust-based fraud, conducted through deepfakes, adaptive language, and emotional manipulation. This year, deception will sound like trust. Enterprises must continuously verify identity, context, and intent across every interaction.

5. Quantum Risk Moves From Long-Term Concern to Near-Term Action

Quantum computing may still be years from cracking today's encryption, but the threat has already changed, and should continue to change, enterprise behavior. Governments, cloud providers, and large enterprises are racing to secure cryptographic agility, migrating from vulnerable Rivest–Shamir–Adleman (RSA) and Elliptic Curve Cryptography (ECC) algorithms to post-quantum cryptography (PQC) standards before adversaries can weaponize them.

The danger lies in the 'harvest now, decrypt later' (HNDL) strategy. Attackers are already stealing encrypted data today, confident that quantum decryption will expose it tomorrow. In 2026, preparation moves from theory to execution. Boards will fund cryptographic bills of materials (CBOMs) to catalogue every algorithm, certificate, and key across their environments. Organizations will pilot National Institute of Standards and Technology (NIST)-approved post-quantum algorithms and pressure vendors to show clear migration timelines.

Quantum risk is not about tomorrow's machines. It is about today's data. Every organization must assume its encrypted assets are already being harvested and prepare for a world where prevention depends on cryptographic agility.

6. AI Becomes a Strategic Decision Engine

AI is steadily changing the foundations of cyber security. What once served mainly as a tool for operational efficiency is now influencing how both attackers and defenders plan, adapt, and execute their strategies. The industry is moving into a phase where AI is no longer a supporting capability, but an embedded element in detection, analysis, and decision-making workflows.

This evolution is expected to deepen. Attackers are already using AI to generate faster, broader, and more tailored campaigns, and this will increasingly push organizations to develop defensive capabilities that can match that pace, with continuous learning, real-time context, and more autonomous operational support. It reflects a shift in how security teams prioritize actions, understand risk, coordinate response, and ultimately, increase efficiency.

AI is becoming an integral part of the operational layer within security operations, enhancing human expertise, simplifying manual workflows, and reducing the mean time to remediation (MTTR).

The accelerated adoption of AI is making it part of the operational backbone of cyber security rather than an extension of existing tools, shaping analytical workflows and decision-making processes to be more consistent, automated, and guided by precise controls.

“THE ACCELERATED ADOPTION OF AI IS MAKING IT PART OF THE OPERATIONAL BACKBONE OF CYBER SECURITY RATHER THAN AN EXTENSION OF EXISTING TOOLS, SHAPING ANALYTICAL WORKFLOWS AND DECISION-MAKING PROCESSES TO BE MORE CONSISTENT, AUTOMATED, AND GUIDED BY PRECISE CONTROLS.”

7. The AI Reality Check

After two years of near-frantic AI adoption, we will experience our first major recalibration. Many organizations that rushed to integrate generative AI tools will discover ungoverned systems, exposed APIs, and compliance blind spots. Shadow AI, employee-initiated tools using corporate data, will proliferate, creating invisible data leaks and inconsistent security standards.

This phase of disillusionment is necessary: it will drive the shift from experimentation to accountability. Executives will begin demanding AI value measured in outcomes, not hype. AI assurance frameworks will emerge across various sectors, necessitating formal audits to ensure fairness, robustness, and security. Leadership teams must establish clear policies for AI use and align them with legal, ethical, and risk frameworks. Responsible deployment will hinge on explainability and continuous validation, not unchecked automation. Compliance will expand from privacy to algorithmic accountability.

AI's first disruption was speed; its second will be governance. 2026 will reward those who treat AI not as a shortcut but as a capability to be secured, audited, and improved.

8. Regulation and Accountability Expand - Cyber Resilience Becomes a License to Operate

Regulators worldwide are closing the gap between innovation and accountability. In 2026, regulation ceases to be a reactive approach. Frameworks such as the EU's NIS2 Directive, the AI Act, and the U.S. SEC incident-disclosure rules will converge on a single principle: cyber security must be measurable and demonstrable in real-time. Governments will now expect continuous proof of resilience. Organizations are expected to demonstrate that their preventive controls, incident response plans, and data protection measures are consistently enforced.

There is a strong reason behind this regulatory acceleration: society's growing dependence on digital services to maintain daily life and the economy without significant disruptions. Business resiliency has become the primary driver behind the increasing compliance requirements.

This shift will mark the end of the era of "annual compliance." Enterprises will rely on automated compliance monitoring, machine-readable policies, real-time attestations, and AI-based risk analytics. Boards and CEOs will carry personal responsibility for oversight.

Cyber resilience is no longer paperwork; it's performance. The ability to demonstrate protection continuously will determine market access and trust.

07

2026 CISO RECOMMENDATIONS

BY JONATHAN FISCHBEI
Field CISO

Security leaders' primary challenge in 2026 is maintaining the organization's security as attacker capabilities, techniques, and scale evolve faster than ever before, touching on wider attack surfaces, from everyday workflows and endpoints to hybrid environments composed of increasingly complex webs of systems. At the same time, CISOs are expected to demonstrate, clearly and continuously, operational efficiency and support measurable business outcomes. The recommendations that follow reflect the priorities that CISOs must focus on, including reducing exposure, governing risk in dynamic environments, and demonstrating resilience against an increasingly aggressive and unpredictable threat landscape.

1. Establish Prevention-Led, Layered Security Programs

Security programs must be designed to stop attacks as early as possible while assuming no single preventive control will be sufficient on its own. In 2026, effective programs prioritize prevention at multiple points across the attack chain, reducing exposure and attack success rates, while ensuring secondary, complementary safeguards can contain the impact when prevention is bypassed. This approach moves beyond monolithic defenses and toward layered, adaptive protection that reflects how attackers operate.

CISOs should reinforce prevention-led architectures with continuous validation and transparency mechanisms that confirm protections are working under real conditions. This includes integrating external signals such as responsible vulnerability disclosure, targeted security awareness tied to observed threat activity, and structured programs that surface weaknesses before adversaries exploit them. These elements are not substitutes for prevention, but independent checks that strengthen confidence in defensive effectiveness and accelerate improvement.

Why it matters: Adversaries operate at scale, iterate rapidly, and exploit the first viable weakness they encounter. Organizations that rely on single controls or static assurance models are more likely to experience cascading failure, while prevention-led, layered programs reduce both the likelihood and impact of successful attacks.

2. Govern Data Protection as a Core Security Outcome

Data exposure now represents the most consequential outcome of modern cyber incidents, exceeding the business impact of service disruption alone. Security programs must therefore treat data protection as a first-order objective, governed by how sensitive data is accessed, moved, and aggregated across environments—not by static classifications or perimeter assumptions. In this context, ransomware incidents should be addressed by default as data-exposure events, with availability loss viewed as only one dimension of impact.

CISOs should prioritize architectural controls that limit data blast radius and recovery risk, including segmentation of data access paths, strict least-privilege enforcement, and resilience measures such as immutable backups and regularly exercised incident-response playbooks. These controls must be designed to assume partial compromise and focus on preventing large-scale data exfiltration, accelerating recovery, and preserving trust during and after an incident. This governance model also establishes a foundation for emerging risk domains, including the long-term protection of sensitive data against future advancements in cryptography.

Why it matters: Ransomware campaigns increasingly involve confirmed data exfiltration, making data exposure, rather than downtime, the primary source of regulatory, financial, and reputational impact. Organizations that fail to govern data protection as a core security outcome remain vulnerable to compounding loss during incidents and long-tail risk as data persists beyond today's threat horizon.

3. Operationalize Cloud, SaaS, and AI Security

Cloud, SaaS, and AI environments introduce risk primarily through speed, scale, and change, not just misconfiguration. Continuous deployment, third-party integrations, and automated service interactions create exposure that cannot be effectively governed through identity-centric or compliance-driven controls alone. These platforms must be secured as living, operational systems, where risk emerges from how services interact and execute in real-time.

CISOs should establish governance that continuously evaluates platform posture and operational behavior, including configuration drift, API usage, service-to-service trust, and application-level interactions. AI systems require the same operational discipline as other production platforms, with defined ownership, monitored usage boundaries, and accountability for how models are accessed, integrated, and acted upon. The focus is not on measuring resilience or enforcing identity policy, but on maintaining ongoing control over how dynamic platforms behave as they evolve.

Why it matters: Attackers increasingly exploit APIs, automation, and runtime interactions to bypass identity checks and perimeter defenses.

Without continuous platform governance, organizations lose visibility and control between change cycles, creating exploitable gaps that scale as environments become more complex.

4. Treat Third-Party Risk as Structural Exposure

Vendors, SaaS providers, and partners are embedded directly into enterprise environments through access, integrations, and shared services.

CISOs should manage third-party risk as structural exposure, not as a periodic assessment exercise. This requires continuous monitoring of vendor access, segmentation of partner connections, and enforcement of least-privilege and Zero Trust principles across external identities.

Security obligations must be measurable and enforceable through SLAs, but documentation alone is insufficient. Real risk emerges from how vendors access systems, what permissions they hold, and how compromise propagates across shared trust relationships.

Why it matters: Supply-chain and SaaS-linked incidents increasingly originate from trusted vendor access and inherited trust, expanding blast radius beyond the organization's direct control.

5. Anchor Zero Trust Architecture in Human and Non-Human Identity

Zero Trust must be treated as a core defense against identity-driven attacks. As phishing, credential theft, and token abuse enable attackers to operate as trusted users and non-human identities, these methods dramatically expand the attack surface, rendering implicit trust models and static access assumptions obsolete. Effective Zero Trust requires continuous verification of identity and context, least-privilege access by default, and architectural controls that limit lateral movement across cloud, SaaS, network, and development environments. The goal is not only to prevent misuse of identity, but to contain the impact when identity is inevitably compromised.

The same Zero Trust principles underpin resilience against AI-enabled social engineering, cloud-driven attack surface expansion, and ransomware operations that increasingly rely on stolen identities rather than exploits.

Why it matters: Zero Trust is the practical framework for managing modern identity risk—reducing implicit trust, constraining blast radius, and ensuring identity compromise does not automatically translate into widespread access or business disruption.

6. Harden Trust-Based Business Processes Against Abuse

Attackers are increasingly monetizing access by exploiting implicit trust in business workflows, rather than relying solely on technical compromise. Business Email Compromise (BEC), executive impersonation, and vendor fraud remain highly effective because they exploit legitimate communication channels to initiate financial transactions, disclose sensitive data, or modify access permissions. With AI-driven impersonation on the rise, including realistic phishing lures, synthetic voice, and deepfake-assisted social engineering, these attacks have further increased their credibility and scale, reducing the time required to move from initial contact to impact.

CISOs should treat trust-based business processes as a core component of the threat surface and apply security controls accordingly. This includes strengthening protections across email and collaboration platforms, implementing mandatory, context-aware verification for high-risk actions, and eliminating single-step approvals and other implicit trust assumptions from payment, vendor, and access-granting workflows. These controls are particularly critical for executives and high-risk business functions, where successful impersonation can enable immediate financial loss or serve as a precursor to ransomware deployment by facilitating credential resets, access expansion, or data exfiltration.

Why it matters: BEC and impersonation attacks are increasingly serving as both direct monetization vectors and enablers for ransomware and extortion campaigns. As attackers blend social engineering, identity abuse, and AI-enabled deception, organizations that fail to harden trust-based workflows remain vulnerable to high-impact outcomes, even in environments with strong perimeter and endpoint defenses.
Top of Form
Bottom of Form

7. Integrate OT and Cyber Risk Governance

Operational Technology (OT) environments now sit at the intersection of cyber risk, physical safety, and business continuity. As OT environments increasingly adopt Industry 4.0 architectures—extending industrial systems through IoT and IIoT devices, cloud connectivity, and remote access— increased connectivity between IT and OT—driven by remote access, cloud monitoring, and digital transformation—has expanded attack paths into environments where compromise can result in physical disruption, safety incidents, or prolonged operational downtime, not just data loss.

Traditional IT security models are insufficient in these settings, where availability and deterministic behavior are paramount, and active security controls must be applied with care. Securing this new, cloud-connected OT environment requires security approaches that are aligned with its expanded digital and operational exposure.

CISOs should ensure that OT security is governed through a risk-based model aligned with operational realities, rather than being treated as a standalone technical domain. This includes enforcing strict and continuously validated IT-OT segmentation, leveraging passive and non-intrusive monitoring to maintain visibility without disrupting operations, and integrating OT telemetry into centralized SOC workflows to enable early detection of abnormal activity.

Just as importantly, cyber, engineering, and physical safety teams must operate from a shared risk framework that prioritizes safety, uptime, and resilience, ensuring cyber incidents in OT environments are assessed and responded to with full awareness of their potential physical and operational impact.

Why it matters: As attackers increasingly target industrial and critical infrastructure environments, cyber incidents in OT no longer represent isolated technical events—they directly translate into safety risks, operational disruptions, and material business impacts. Organizations that fail to secure this modern, Industry 4.0 OT environment as part of enterprise cyber risk governance remain vulnerable to attacks that bypass traditional IT defenses and exploit the gap between cyber and physical security.

8. Prove Resilience, Not Just Compliance

Cyber resilience can no longer be inferred from policy adherence or point-in-time assessments. As attack surfaces change continuously and threats exploit exposure faster than review cycles can detect, resilience must be measurable, continuously validated, and expressed in business-relevant terms. Annual audits and static risk assessments may satisfy regulatory requirements, but they do not reflect an organization's real-world ability to withstand, contain, and recover from active threats.

At the same time, increasing complexity across cloud, automation, and AI environments makes it critical not to lose focus on cyber hygiene and security fundamentals. Many successful attacks still exploit basic weaknesses, underscoring that resilience depends on maintaining strong "back-to-basics" practices alongside newer capabilities.

CISOs should shift toward continuous control validation and exposure-driven measurement, integrating telemetry from across the environment to assess not only whether controls exist, but also whether they are effective under real-world conditions. This includes monitoring exposure trends, remediation velocity, and time-to-contain across attack paths, as well as automating evidence collection to reduce manual compliance overhead. Effectiveness must be communicated differently to boards, regulators, partners, and customers, using outcome-based metrics that demonstrate reduced risk, faster response times, and improved containment, rather than merely checklist completion.

Signals from digital trust and transparency programs, such as external vulnerability reporting, third-party findings, and responsiveness to disclosed risk, should be treated as indicators of operational resilience, not reputational liabilities. These signals provide independent validation of how quickly and effectively an organization identifies and addresses exposure.

Why it matters: In an environment of continuous threat activity and increasing regulatory scrutiny, organizations that can demonstrate resilience through ongoing measurement and real-world outcomes are better positioned to maintain trust, meet regulatory expectations, and respond credibly to incidents than those relying solely on periodic compliance assessments.

08

AN EXPOSURE MANAGEMENT PERSPECTIVE

THREAT INTELLIGENCE BEFORE THE BREACH

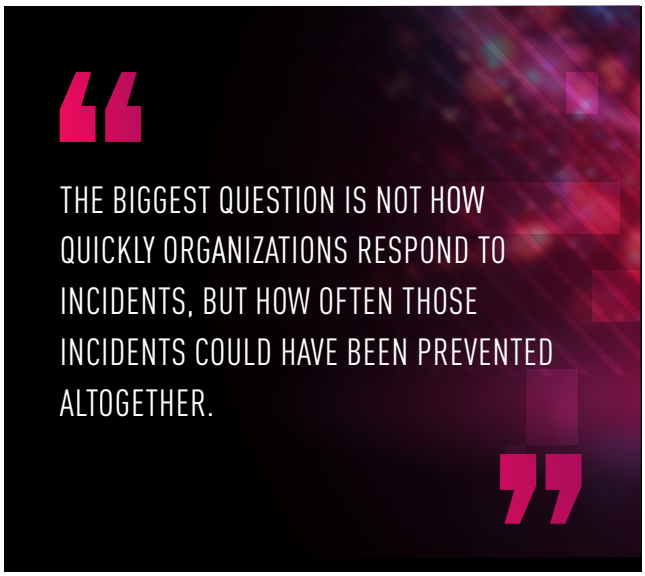
An Exposure Management Perspective

From Incident Response to Preemptive Security

Incident response is a critical function in any security program, but ultimately, it only represents the final stage of most attacks. By the time response teams are engaged, adversaries have already completed reconnaissance, established infrastructure, and initiated access. The biggest question is not how quickly organizations respond to incidents, but how often those incidents could have been prevented altogether.

A growing body of intelligence shows that many attacks leave detectable external signals well before internal compromise occurs. These insights, when understood in context, present an opportunity to reduce reliance on emergency response by addressing exposure earlier in the attack lifecycle. From an exposure management perspective, threat intelligence plays a foundational role in taking security programs from reactive containment to preemptive risk reduction.

This perspective does not replace incident response or post-breach investigation. Rather, it complements them by focusing on the conditions and activities that precede incidents, with the aim of preventing issues from ever reaching the point where a response is required.



What Attackers Do Before the Breach

Before an incident unfolds internally, attackers typically invest time in preparation. This preparatory phase often includes activities that are external to the organization but directly relevant to its security posture. Common examples include the creation of look-alike domains, brand impersonation across social platforms, deployment of phishing infrastructure, harvesting of credentials from prior breaches, and reconnaissance of exposed services. Individually, these events may appear disconnected or benign. However, collectively, they form the earliest indicators of an impending intrusion attempt.

Crucially, these activities surface outside traditional internal monitoring controls. They precede malware execution, lateral movement, or privilege escalation, and therefore occur before most incident response triggers are activated. As a result, they are frequently overlooked or deprioritized, even though they represent the earliest stage at which intervention is possible.

Pre-Incident Intelligence Across the Global Attack Surface

Across the global attack surface, certain patterns consistently emerge during the pre-incident phase. Attackers rarely move directly from intent to exploitation. Traditionally, they stage infrastructure, test delivery mechanisms, and refine targeting based on observed responses.

While the specific prevalence of phishing, impersonation, or credential-based techniques varies by sector and geography, these vectors continue to dominate early attack activity. Importantly, the same preparation techniques observed externally often align closely with the tactics incident response teams later observe post-breach. The challenge for defenders is not the absence of early signals, but the difficulty of recognizing which exposures are relevant, credible, and actionable for a specific organization.

When attacker preparation activity is addressed during this early phase, the downstream impact can be significant. Disrupting phishing infrastructure, neutralizing impersonation assets, or mitigating exposed entry points before exploitation can prevent campaigns from progressing to internal compromise.

From an exposure management standpoint, the goal is not to predict every attack, but to reduce the number of exposures that attackers have available to them. By addressing threats while activity remains external, organizations can alter attack paths before they generate alerts, incidents, or business disruption.

Effective pre-breach disruption often results in the absence of incidents rather than visible response metrics. Attacks fail silently. Users are never impacted. Incident response teams are never engaged. Over time, this reduction in incident volume is one of the clearest indications that exposure is being managed effectively.

Case Patterns: Preempting the Same Threats Incident Response Sees

During incident response investigations, certain attack sequences recur frequently. When viewed retrospectively, many of these incidents follow a progression that was externally visible before internal compromise occurred.

EXTERNAL IMPERSONATION → INTERNAL CREDENTIAL THEFT

In numerous cases, attacker activity begins with external brand impersonation, such as look-alike domains, spoofed communications, or fraudulent social profiles, all designed to establish credibility. These assets are then used to harvest credentials, which later become the primary mechanism for internal access. By the time credential abuse is detected internally, the initial impersonation infrastructure has often already served its purpose.

PHISHING INFRASTRUCTURE → ESCALATION TO INCIDENT RESPONSE

Phishing campaigns that ultimately result in incident response engagement rarely appear without warning. The infrastructure supporting these campaigns typically surfaces in advance. When this infrastructure remains active long enough to reach users, the likelihood of escalation increases significantly. Alternatively, when such infrastructure is disrupted early, many campaigns fail to progress beyond initial delivery attempts, never reaching their potential.

EXPOSURE INTELLIGENCE → COMPENSATING CONTROLS BEFORE EXPLOITATION

Not all pre-breach activities are identity-driven. In many cases, attacker preparation aligns with known weaknesses in exposed services or configurations. When intelligence indicates active interest in specific attack paths, organizations can apply compensating controls such as temporary mitigations or virtual protections to reduce exposure while permanent remediation is pending. These remediations can break exploitation chains before attackers move from reconnaissance to execution. already served its purpose.

Proactive Defense Insights

- Attacker preparation will continue to accelerate, driven by automation and readily available infrastructure.
- The likelihood of breach will correlate less with severity scores and more with exposure duration and attacker readiness.
- Organizations that align threat intelligence with exposure reduction will reduce incident frequency over time.

Closing the Gap: Threat Intelligence as the First Step of Exposure Management

Incident response provides invaluable insight into how attacks succeed. Exposure management applies those lessons earlier in the lifecycle, utilizing threat intelligence to identify and mitigate risk before incidents occur.

By connecting pre-incident intelligence with post-incident learnings, organizations can close the gap between what they respond to and what they prevent. Threat intelligence is not an add-on or a feed; it is the starting point for understanding exposure, prioritizing action, and reducing the volume of incidents that require a response.

> Incident response shows us how attacks succeed. The real opportunity exposure management provides is threat intelligence - using those lessons earlier, when attacker activity is still external and exposure management can stop incidents before response is ever required.

MICHAEL GREENBERG

Head of Product Marketing
Exposure Management

ABOUT CHECK POINT SOFTWARE TECHNOLOGIES LTD.

Check Point Software Technologies Ltd. (www.checkpoint.com) is a leading AI-powered, cloud-delivered cyber security platform provider protecting over 100,000 organizations worldwide. Check Point leverages the power of AI everywhere to enhance cyber security efficiency and accuracy through its Infinity Platform, with industry-leading catch rates enabling proactive threat anticipation and smarter, faster response times. The comprehensive platform includes cloud-delivered technologies consisting of Check Point Harmony to secure the workspace, Check Point CloudGuard to secure the cloud, Check Point Quantum to secure the network, and Check Point Infinity Core Services for collaborative security operations and services.

CONTACT US

WORLDWIDE HEADQUARTERS

5 Shlomo Kaplan Street, Tel Aviv 6789159, Israel

Tel: 972-3-753-4599

Email: info@checkpoint.com

U.S. HEADQUARTERS

100 Oracle Parkway, Suite 800, Redwood City, CA 94065

Tel: 800-429-4391

UNDER ATTACK?

Contact our Incident Response Team:

emergency-response@checkpoint.com

CHECK POINT RESEARCH

To get our latest research and other exclusive content,

Visit us at www.research.checkpoint.com

www.checkpoint.com

© 2026 Check Point Software Technologies Ltd. All rights reserved.