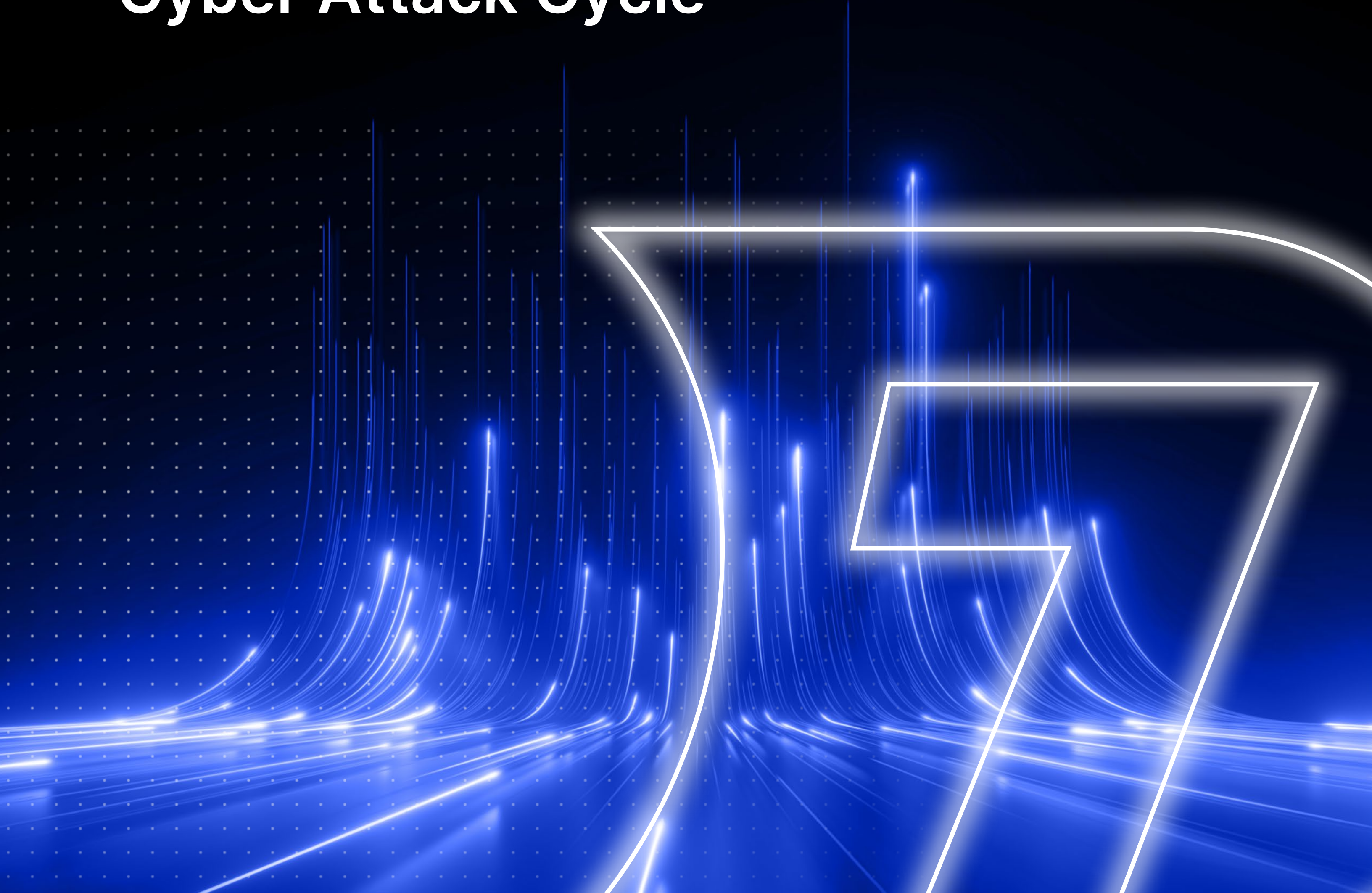
RAPID7

2026 GLOBAL THREAT LANDSCAPE REPORT

Decoding the Accelerated Cyber Attack Cycle



CONTENTS

INTRODUCTION

The role of intelligence is evolving, as threat actors gather greater access to tools that compromise organizations globally and geopolitics continues its transition to becoming either entirely digital or hybrid. Our 2026 report demonstrates that cyber intelligence is now more important than ever, in particular, the need for actionable insights. For example, the growth in the number of exploited vulnerabilities means remediation has to keep pace with these changes. In other words, reviewing security advisories, determining exposure, and implementing updates to manage the risk each must keep pace with well-funded threat groups.

This demonstrates exactly why vulnerability intelligence is quickly becoming a critical component of an organization’s threat intelligence strategy. Understanding these initial entry vectors, used by threat groups that previously relied almost entirely on weak or compromised credentials, is imperative given that we know the number of criminal groups are increasing (e.g., ransomware as detailed within the report).

Traditional threat intelligence, which focused on collating large datasets of Indicators of Compromise (IoCs), is broadening to demand higher efficacy and context. This new, broader perspective involves integrating preemptive indicators, such as dark web signals, into alerts to accurately identify potential exposure before an attack materializes. Furthermore, it emphasizes the importance of prioritizing the remediation process to address exposures before adversaries gain a network foothold.

Rapid7 Labs embodies this approach, using "curated intelligence" and "actionable insights" derived from a broad ingestion of signals to provide accurate, actionable information for both organizations and the open-source community.

— The Rapid7 Labs Team

EXECUTIVE SUMMARY

Cyber risk fundamentally changed in 2025. For years, organizations invested in faster detection and response, assuming that speed would offset exposure. That assumption no longer holds. The balance has shifted from reacting quickly to anticipating risk before it materializes.

The core drivers of cyber incidents remain unchanged. Weak credentials, unpatched systems, and exposed services continue to account for the majority of successful intrusions. What has changed is the pace at which these conditions are identified and exploited. Automation and artificial intelligence (AI) have dramatically compressed the time between exposure and impact, often reducing the opportunity for meaningful intervention to minutes, or eliminating it altogether.

This report does not suggest a sudden transformation in attacker intent or sophistication. Rather, it reflects the acceleration of existing methods. AI is being used to scale reconnaissance, automate decision-making, and industrialize social engineering, compressing the time between exposure and exploitation. Our findings show that the majority of successful intrusions still originate from known, preventable conditions: exposed services, weak identity controls, and unpatched edge infrastructure. What has changed is how quickly those conditions are discovered and weaponized.

> “Security did not fail in 2025 because defenders were slow. It failed because speed was no longer the advantage.”

This shift requires a fundamental mindshift toward preemptive security. Preemptive security means reducing the conditions attackers rely on before exploitation occurs, detecting and responding with full environmental context, and prioritizing action based on material risk, not alert volume. Organizations that fail to adopt this approach face a widening asymmetry: as attacker velocity increases, reactive decision models become increasingly misaligned with how risk now materializes.

At the same time, the traditional perimeter of the enterprise has dissolved. Our findings show attackers consistently targeting the most trusted and operationally critical layers of modern organizations, including identity systems, cloud environments, and collaboration platforms. These environments blur the distinction between legitimate activity and malicious behavior, increasing the difficulty and cost of containment once access is established.

In this environment, the advantage comes from clarity. Organizations that continuously understand their attack surface, apply curated intelligence to what matters most, and connect technical exposure to business impact are best positioned to reduce risk before it becomes disruption. In 2026, effective cyber defense is defined by informed prioritization and anticipation rather than simply reacting to alerts.

KEY FINDINGS

Cyber risk did not transform in 2025 because attackers discovered entirely new techniques. It transformed because the entire ecosystem surrounding compromise accelerated, from underground access markets, to ransomware deployment, to nation-state pre-positioning. What once unfolded over weeks now materializes in days, and in some cases, minutes.

KEY FINDING #1: THE PREDICTIVE WINDOW HAS COLLAPSED

In 2025, the statistical buffer between disclosure and exploitation materially narrowed:

- Within newly disclosed CVSS 7–10 vulnerabilities, confirmed exploitation increased 105% year over year, rising from 71 in 2024 to 146 in 2025.
- Median time from publication to CISA KEV inclusion dropped from 8.5 days to 5.0 days, while mean time dropped from 61.0 days to 28.5 days.
- The number of “high-risk but not yet exploited” vulnerabilities (EPSS ≥ 0.7 without confirmed exploitation) fell dramatically, indicating that high-probability vulnerabilities are being operationalized almost immediately.

KEY FINDING #2: ATTACKERS ARE MONETIZING UNMANAGED EXPOSURE

The intrusion lifecycle increasingly begins with known, preventable exposure conditions rather than breakthrough exploitation techniques:

- Valid account / no MFA accounted for 43.9% of all IR incidents in 2025, making it the single most common initial access vector.
- Vulnerability exploitation accounted for 24.6%, and exposed services for 7.0%.
- On underground forums, RDP (21.2%), VPN (12.8%), and RDWeb (11.2%) were the most frequently advertised access types.
- “Domain User” privileges were the most commonly sold level of access.

KEY FINDING #3: EXPLOITATION IS CONCENTRATING ON A NARROW SET OF RELIABLE WEAKNESS CLASSES

Rapid7's Vulnerability Intelligence team observed that confirmed exploitation clustered around a small number of weakness classes:

- CWE-502 (Deserialization) was the most common root cause among exploited vulnerabilities.
- Authentication bypass and memory corruption vulnerabilities remained consistently represented in confirmed exploitation data.
- Several high-profile ransomware campaigns focused on deserialization flaws and authentication bypasses in file transfer systems, edge appliances, and collaboration platforms.

KEY FINDING #4: RANSOMWARE HAS MATURED INTO A SPEED-OPTIMIZED ACCESS ECONOMY

Ransomware was not a peripheral threat in 2025, it was the dominant operational outcome. 42% of Rapid7 MDR incident response investigations in 2025 involved ransomware.

At the same time:

- Total ransomware leak posts increased from 6,034 in 2024 to 8,835 in 2025 (a 46.4% YoY rise).
- The number of unique active ransomware groups grew from 102 to 140.
- Data theft increasingly preceded encryption, reinforcing smash-and-grab extortion models.

KEY FINDING #5: EMBEDDED ACCESS, NOT PERIMETER BREACH, DEFINES STRATEGIC RISK

Across both financially motivated and state-aligned operations, adversaries converged on the same high-value control surfaces:

- Telecommunications and network-edge infrastructure
- Cloud identity and device-code authentication flows
- Collaboration platforms abused as command-and-control channels
- SaaS APIs and trusted third-party integrations

THE DISAPPEARANCE OF PREDICTIVE LEAD TIME

If we take a careful look at the CVE data coming out of 2025, the story isn't really about volume anymore. Yes, vulnerability counts continue to rise, and yes, the numbers are large enough to overwhelm most programs on their own. But that problem is familiar. What stands out in the 2025 data is something more subtle and, frankly, more difficult to deal with: risk is no longer accumulating quietly. It is being realized almost immediately.

That shift changes how we should interpret nearly every metric we rely on today.

Volume increased and exploitation accelerated

At a high level, the growth in high and critical vulnerabilities (CVSS 7–10) from 2024 to 2025 was notable but not unprecedented. The total count increased from roughly 16,200 in 2024 to just over 18,100 in 2025, which is consistent with longer-term disclosure trends rather than a dramatic inflection point on its own.

The more meaningful change appears when looking specifically at exploitation within the high to critical CVSS constrained dataset. For this analysis, "exploited in the wild" refers only to vulnerabilities that meet all three of the following criteria: a CVSS score between 7 and 10, publication in the same calendar year being analyzed, and confirmed exploitation within that dataset. Using that scope, the number of exploited vulnerabilities increased by approximately 105% year over year, rising from 71 in 2024 to 146 in 2025 (see Figure 1). This increase far outpaced the year-over-year growth in newly disclosed vulnerabilities within the CVSS 7–10 range. As shown in the figure below, this acceleration is not a single-year anomaly.

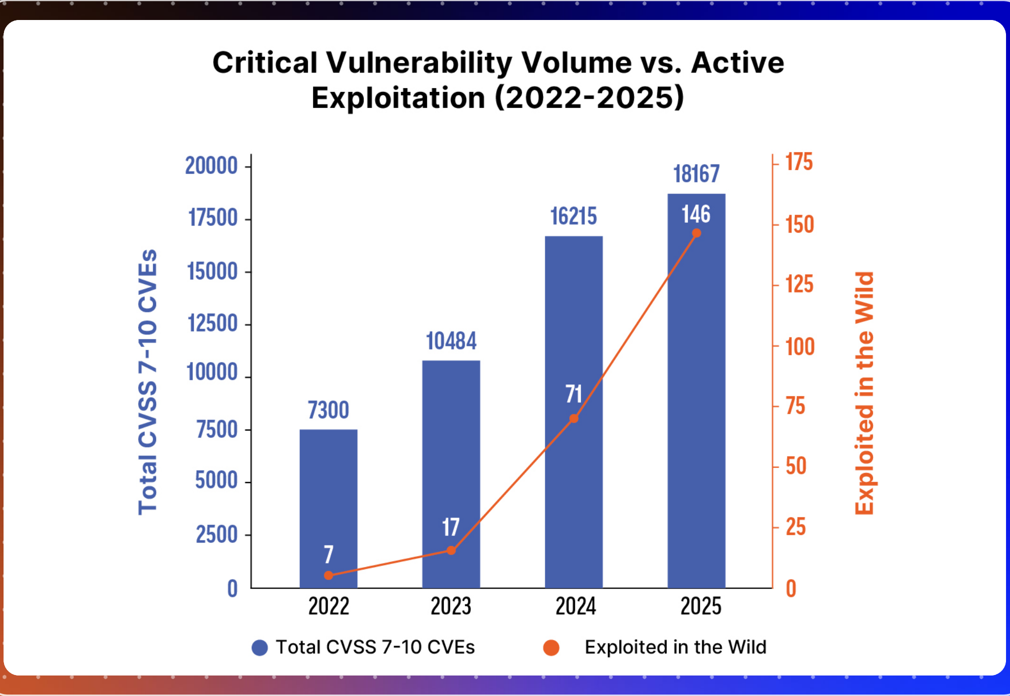


Figure 1

Across the four-year window from 2022 to 2025, exploited vulnerabilities grew roughly twentyfold, from 7 to 146, while the total number of CVSS 7–10 disclosures more than doubled. The exploitation rate relative to disclosure volume has shifted from less than 0.1% in 2022 to 0.8% in 2025. These combined findings indicate that attackers are weaponizing a larger share of the critical vulnerability surface each year. The growing volume of critical disclosures is not merely expanding the attack surface, but actively outpacing defenders' capacity to prioritize and patch, creating an environment where a larger proportion of available vulnerabilities are successfully weaponized each year.

Additional timing data reinforces the trend toward faster exploitation. When restricting the analysis to CVSS 7–10 vulnerabilities published in the same calendar year and later added to CISA KEV, the median time from publication to KEV inclusion dropped from 8.5 days in 2024 to 5.0 days in 2025, with the mean dropping from 61.0 days to 28.5 days. While not every vulnerability is exploited immediately, the overall distribution has shifted noticeably toward shorter timelines.

Taken together, these findings suggest that attackers are not simply benefiting from an expanding pool of vulnerabilities. They are becoming more efficient at identifying which newly disclosed, high-severity vulnerabilities are worth operationalizing and doing so more quickly. From a defensive perspective, this further compresses the decision window. The challenge is increasingly not whether a vulnerable system will eventually be patched, but whether exposure can be identified and addressed before exploitation becomes widespread.

While AI has a part to play in this acceleration, it is further fueled by the industrialization of the cybercrime ecosystem, where Initial Access Brokers (IABs) and specialized collectives remove operational friction by selling verified access and pooling expertise. Attackers have strategically pivoted toward abusing valid credentials and converging on edge infrastructure to bypass hardened perimeters. This shift, combined with "smash-and-grab" tactics that prioritize immediate exfiltration, ensures that risk is realized almost immediately after a vulnerability is operationalized.

CWE shifts reveal attacker preferences

Before looking at how CWEs shifted year over year, it is important to separate two related but distinct views of the data: CWEs ranked by total CVE volume and CWEs associated with vulnerabilities actually exploited in the wild. These two perspectives often diverge and, in 2025, that divergence became more pronounced.

From a volume standpoint, the overall picture remained relatively stable at the top. CWE-79 (Cross-Site Scripting) and CWE-89 (SQL Injection) continued to be the most common weaknesses disclosed. However, the broader top-five composition shifted meaningfully from 2024. Memory handling issues such as CWE-119, filename control problems like CWE-98, and CWE-352 (Cross-Site Request Forgery) re-emerged prominently in disclosure counts. This reflects ongoing software complexity and recurring development patterns rather than a sudden change in attacker behavior.

That picture changes when the focus shifts to vulnerabilities exploited in the wild. In 2025, CWE-502 (Deserialization) was the most common root cause among exploited vulnerabilities, followed by memory corruption issues and command injection. Notably, several CWEs that ranked highly by disclosure volume, including CSRF, were far less prominent in confirmed exploitation data.

This distinction matters. High-volume CWEs largely reflect how software is built. Exploited CWEs reflect how attackers operate. The two are related, but they are not interchangeable. Counting vulnerabilities tells us where defects exist. Tracking exploited weaknesses tells us where attackers consistently succeed.

Across multiple years, attackers have continued to favor weaknesses that offer reliability and scale. These are issues that enable pre-authentication access, remote code execution, or rapid data exfiltration. The persistent presence of deserialization flaws among exploited vulnerabilities helps explain why file transfer systems, management platforms, and edge-facing services remain disproportionately attractive targets, even when those weaknesses are not the most common by raw CVE count.

EPSS is declining, but that doesn't mean risk is

One of the more counterintuitive trends in the 2025 data is the decline in average EPSS (Exploit Prediction Scoring System) scores. Across the full population of critical vulnerabilities, the average EPSS score dropped significantly compared to 2024 (see Figure 2).

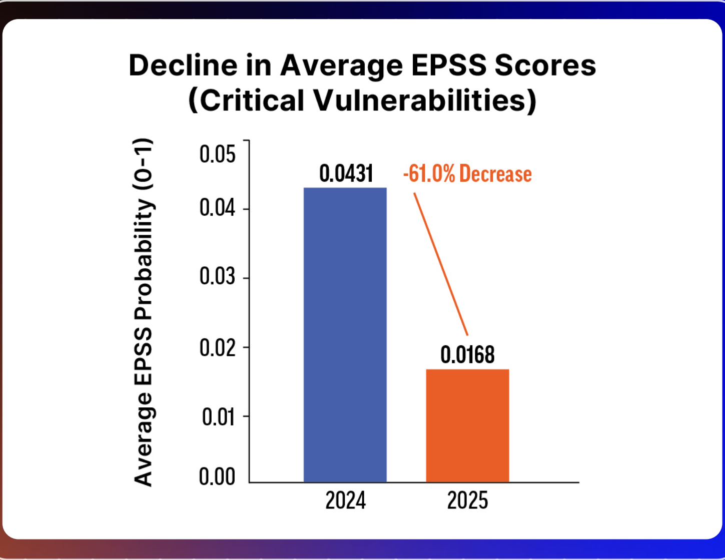


Figure 2

At first glance, this might suggest improvement: fewer vulnerabilities appear likely to be exploited. In isolation, that interpretation is reasonable. EPSS may indeed be getting better at identifying that the vast majority of CVEs never become operational attack vectors.

The problem is what happens at the extremes.

Despite lower average EPSS scores, the number of exploited vulnerabilities increased sharply, jumping from 71 in 2024 to 146 in 2025 in critical vulnerabilities (CVSS 7–10). Even within subsets such as vulnerabilities already confirmed as exploited in the wild, average EPSS scores declined year over year. The same pattern appears in the data for Emerging Threat Response (ETR) vulnerabilities.

This essentially creates a paradox: predicted probability is going down, while realized exploitation is going up (see Figure 3).

The Paradox: Rising Exploitation vs. Declining Predictive Scores

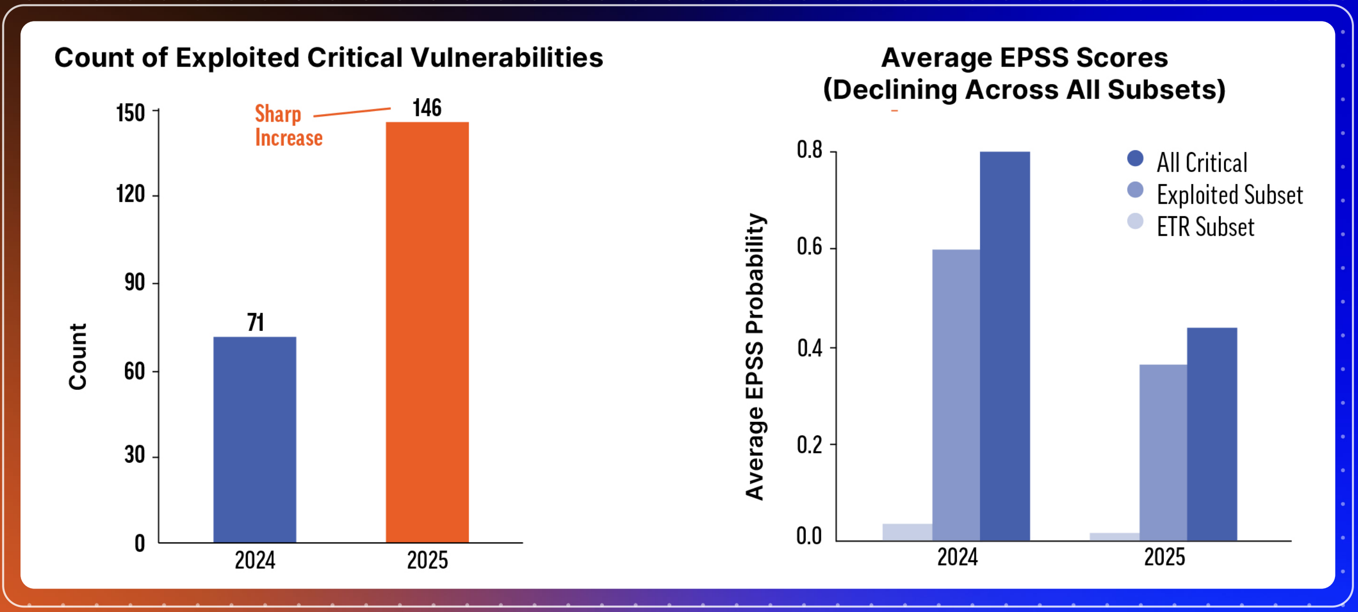


Figure 3

The most plausible explanation is not that EPSS is “wrong,” but that the window in which prediction is useful has narrowed. Vulnerabilities no longer remain in a high-probability, pre-exploitation state for very long. Once a vulnerability is identified as attractive, it often transitions directly into active exploitation.

This shift is most clearly visible in the collapse of the “high-risk but unexploited” category of vulnerabilities. In 2024, there were 338 CVEs with an EPSS score of 0.7 or higher that had not yet been observed exploited in the wild. These vulnerabilities represented latent risk — flaws that defenders knew were likely to be weaponized but still had some measurable window for prioritization, mitigation, or compensating controls.

In 2025, that number fell to just 65, despite an overall increase in both disclosed vulnerabilities and confirmed exploitation. The same pattern appears even more starkly in the Emerging Threat Response (ETR) dataset. In 2024, 15 ETR vulnerabilities fell into the high-risk-but-unexploited category. In 2025, that number dropped to one, while the number of ETR vulnerabilities confirmed as exploited in the wild surged from 5 to 22.

The implication is not that fewer high-risk vulnerabilities exist. Rather, it suggests that high-risk vulnerabilities are no longer remaining unexploited long enough to be managed as latent risk. Once a vulnerability is identified as highly exploitable — either through EPSS, ETR designation, or public disclosure — it is being operationalized by attackers almost immediately. This aligns with broader industry observations that the time between disclosure and confirmed exploitation has continued to compress, often to a matter of days or less.

The key takeaway is that defenders are losing the buffer between “this is likely to be exploited” and “this is actively being exploited.” In prior years, predictive indicators helped buy time. In 2025, that predictive window has largely collapsed. Risk is no longer something that accumulates quietly; for the highest-impact vulnerabilities, it is being exercised almost as soon as it becomes visible.

From predictive signal to near-certain outcome

The Emergent Threat Response (ETR) data makes this trend especially clear. Rapid7’s ETR program is designed to highlight a small set of vulnerabilities that combine high severity, credible exploitation pathways, and meaningful impact.

In 2024, only a minority of ETR vulnerabilities were confirmed as exploited in the wild. In 2025, nearly all of them were.

The volume of ETR-designated CVEs remained stable year over year, and their average CVSS scores stayed extremely high. What changed was exploitation. The category effectively shifted from “likely to be exploited” to “expected to be exploited.”

This change has important implications. It suggests that once a vulnerability reaches a certain threshold of visibility and impact, exploitation is no longer a probabilistic outcome, it is an operational assumption.

Smash-and-grab and the compression of response time

Overlaying all of this is the continued maturation of “smash-and-grab” extortion campaigns. Earlier research documented the initial shift toward these operations, where attackers prioritized speed over persistence and moved from initial compromise to data exfiltration in hours or even minutes. The 2025 data shows that this behavior was not a short-lived tactic but a durable operating model. What was emerging in prior years has now become routine, with attackers increasingly optimizing for rapid exploitation, immediate access to sensitive data, and minimal dwell time rather than long-term footholds.

The specific vulnerabilities weaponized by ransomware campaigns in 2025 validate this operational shift, showing a ruthless concentration on file transfer logistics and network perimeters. Data from 2025 confirms ransomware focusing on CVEs such as CVE-2025-10035 in Fortra GoAnywhere MFT and CVE-2025-49704 in Microsoft SharePoint, both of which allowed attackers to bypass defenses and exfiltrate sensitive enterprise data rapidly. Simultaneously, the perimeter remained a primary battleground as groups leveraged CVE-2025-0282 in Ivanti Connect Secure and CVE-2025-5777 in Citrix NetScaler as reliable initial access vectors. Notably, the root causes driving these ransomware-enabling flaws, primarily deserialization (CWE-502) and authentication bypass, mirror the broader exploitation trends observed this year. This confirms that ransomware operators are prioritizing the speed and reliability of these specific weakness classes to maximize extortion pressure.

This tactic aligns closely with the observed CWE patterns and ETR exploitation rates. It also explains why traditional assumptions about dwell time, phased intrusion models, and gradual escalation are becoming less reliable.

From a purely analytical standpoint, this forces a shift in how risk should be discussed. Vulnerability management alone does not capture exposure. Exposure depends on asset role, network placement, accessibility, and how quickly anomalous behavior can be observed once exploitation begins.

This is where concepts like exposure management and managed detection and response become relevant, not as products, but as responses to measurable changes in attacker behavior. The data suggests that understanding where a vulnerability exists and how quickly exploitation is detected now matters as much as the vulnerability itself.

What the 2024–2025 comparison really tells us

Taking a step back, the comparison between 2024 and 2025 points to a clear conclusion: the threat landscape is not just expanding, it is accelerating. Predictive indicators are losing lead time. Exploitation is becoming the default outcome for high-impact vulnerabilities. And attackers continue to concentrate on a narrow set of reliable weakness classes.

None of this suggests that any single metric or control has failed. Instead, it highlights a growing mismatch between the speed of modern exploitation and defensive processes that were designed for slower cycles.

The challenge moving forward is less about identifying every vulnerability and more about understanding exposure, prioritizing realistically, and responding within increasingly compressed timelines. While the data does not suggest an easy fix, it clearly indicates that both delayed response and misinformed prioritization have become increasingly costly.

THE INDUSTRIALIZATION OF ACCESS

Rapid7's incident response data and underground marketplace monitoring tell the same story from opposite ends of the intrusion lifecycle.

The access types most frequently observed in our 2025 investigations are the same access types openly advertised for sale on underground forums, reflecting a mature ecosystem in which access is systematically harvested, packaged, priced, and resold to ransomware affiliates and other threat actors.

Threat actors have therefore industrialized access in a way that makes it no longer a byproduct of intrusion, but a commodity.

Initial access vectors and trends

Incident response in 2025, as a whole, was dominated by specific initial access vectors and malware infections. Missing or lax multi-factor authentication (MFA) accounted for 43.9% of all the incidents our team observed last year, with vulnerability exploitation sitting in second place with 24.6%. Third place goes to exposed services (7.0%), with fourth (5.3%) shared by brute force, SEO poisoning, and social engineering.

While the downward trend in Figure 4 may appear positive at first glance, the Y axis is sobering. That's because the MFA-related access vector took the top spot across nine of the previous twelve quarters (only occasionally coming second to vulnerability exploitation). Year-over-year data for 2024 to 2025 reveals a steady decline across most quarters, generally, and a final drop in Q4 2025 resulting in a total for the last four quarters down 7.4% YoY.

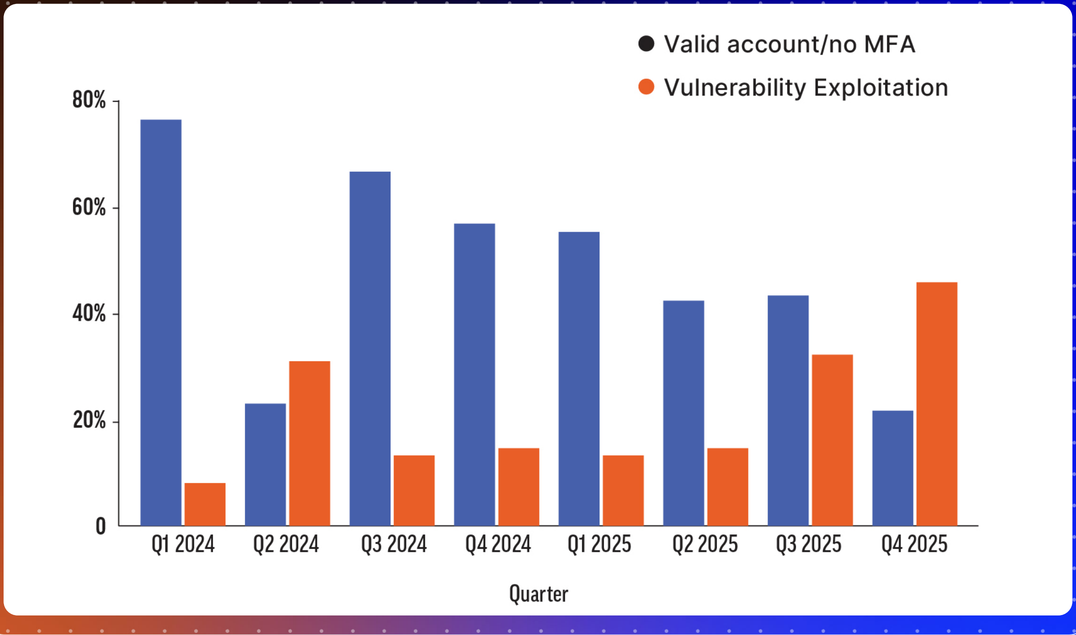


Figure 4

A deeper look at the year highlights additional important themes:

- **Perimeter Weakness:** The exploitation of remote access services (VPNs, RDP) remained the most consistent initial access vector throughout the year, heavily impacting SonicWall, Cisco, and FortiGate appliances.
- **Credential Dominance:** Attackers moved beyond simple theft to sophisticated manipulation, utilizing "Direct Send" features for phishing and social-engineering help desks to bypass MFA.
- **Living Off the Land (LOTL):** Threat actors consistently utilized legitimate administrative tools such as Impacket, Advanced IP Scanner, and WinSCP to mask their lateral movement and exfiltration efforts.
- **Ransomware Evolution:** The threat landscape was dominated by the Akira, Cl0p, and Warlock groups, with a clear trend toward "Data Exfiltration as a Service," where data theft is executed prior to or alongside encryption.
- **MITRE Technique Progression:** Based on the top 10 techniques observed, the landscape shifted from broad account compromise in the early part of 2025 to specific exploitation and impact in Q4 (Figure 5). Whereas the start of the year involved "getting in" via weak credentials and scanning for secrets of repositories and cloud-services, the latter part of the year featured "digging in" by exploiting specific software flaws, manipulating identity controls (i.e., MFA), and executing high-impact ransomware explicitly.

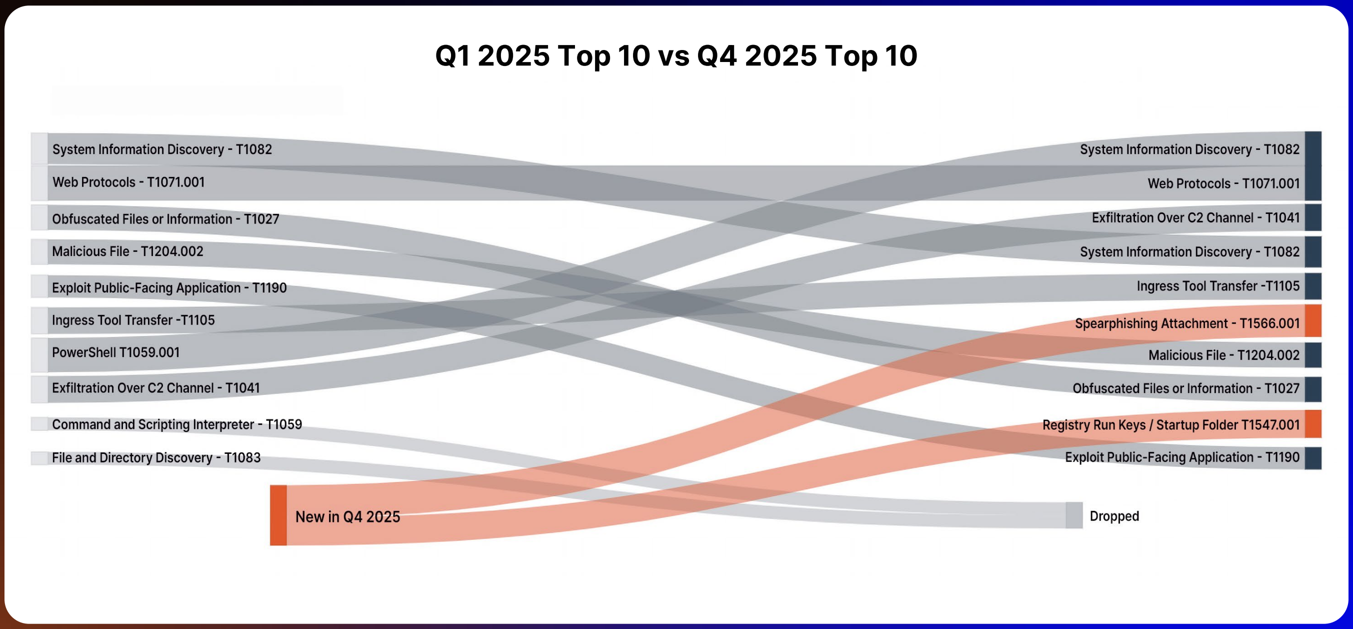


Figure 5

A deeper look at MITRE ATT&CK TTPs

APTs are still very successful in compromising environments with spear phishing emails and attachments, but now they have AI's capabilities to assist them in making it even more convincing. This shows us that the human factor in security risk is important to keep in mind and in check. See Figures 6 and 7, which depict the MITRE ATT&CK tactics, techniques, and procedures (TTPs) our observed threat actor groups are using in their events/attacks, and those that are being used by each industry sector.

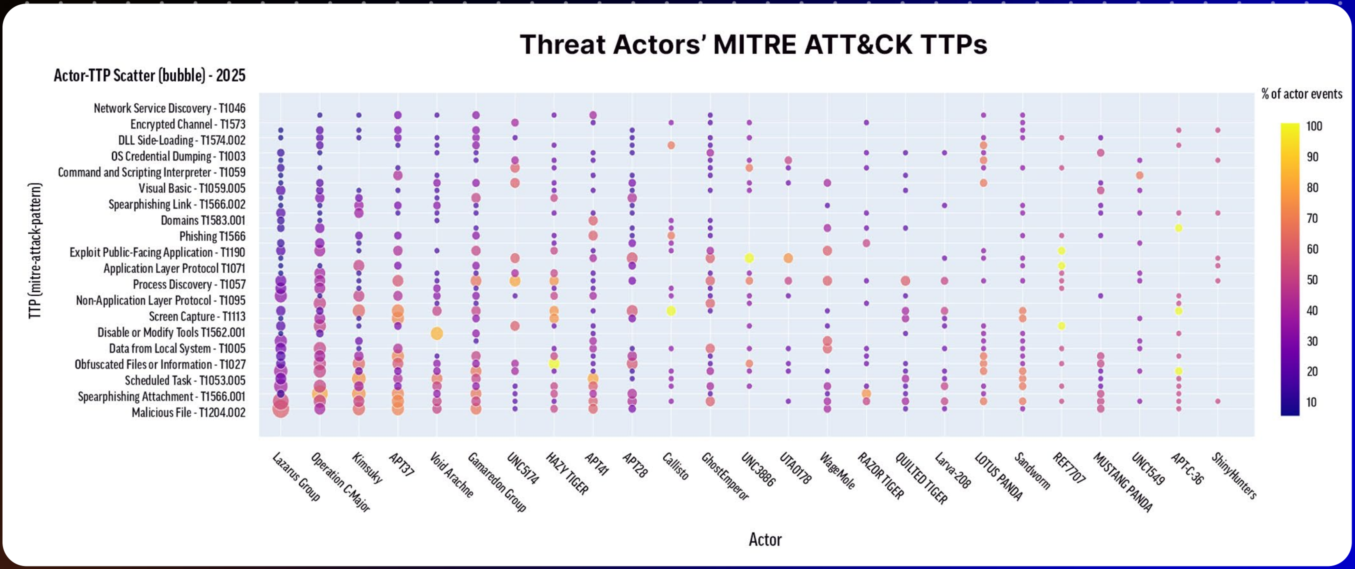

Threat Actors' MITRE ATT&CK TTPs

Figure 6

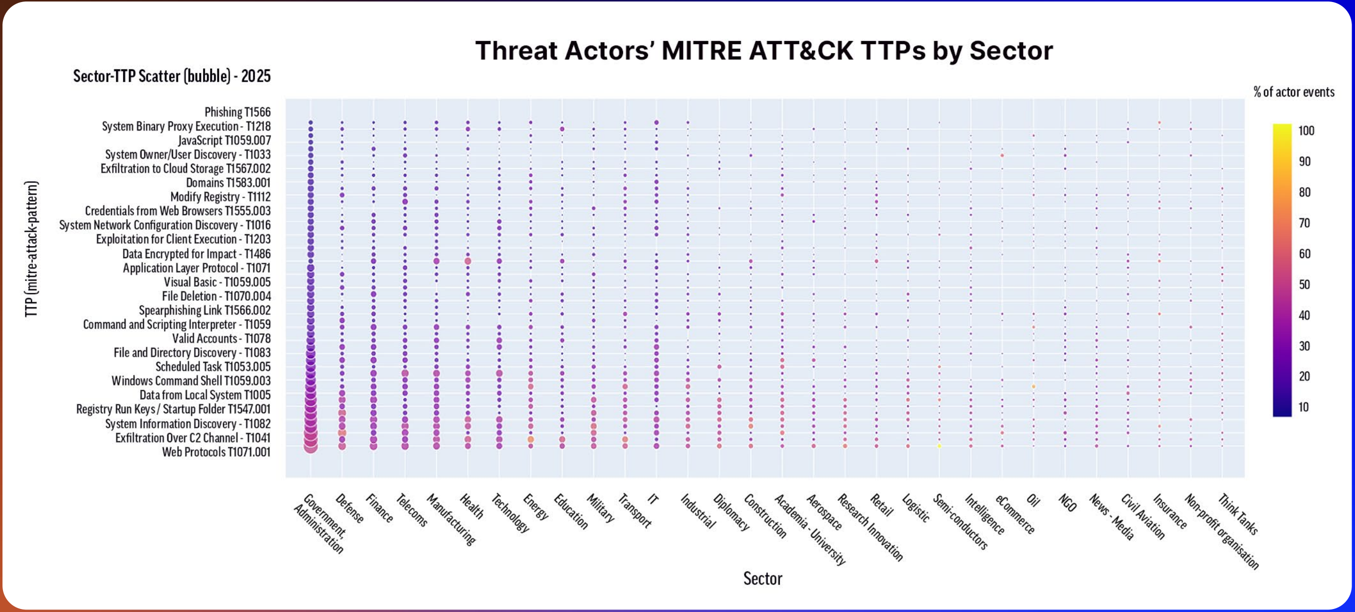

Threat Actors' MITRE ATT&CK TTPs by Sector

Figure 7

Observed malware and exploits

Multiple vulnerabilities were seen to have been exploited across a number of 2025 incident response investigations. These ranged from the targeting of super-admin privileges (CVE-2024-55591, FortiOS and FortiProxy authentication bypass) in Q1, to multiple Microsoft SharePoint vulnerabilities involving authentication, code injection, and code execution in both Q3 and Q4 (CVE-2025-49706, CVE-2025-49704, and CVE-2025-53770).

Elsewhere, vulnerabilities targeting SonicWall SonicOS (CVE-2024-40766), and SonicWall's SSLVPN authentication mechanism (CVE-2024-53704) were observed in investigations across all quarters except Q1.

Malware usage in 2025 solidified trends and numbers we had observed in previous quarters. Malware as a Service (MaaS) phenomenon Bunny Loader accounted for 45.61% of all incidents involving malware. To give a sense of scale, second place belonged to Infostealers with just 12.02% of the overall total. That's not one specific Infostealer — that's all of the Infostealers, including Katz, Filch, Strela, Vidar, and many more.

Threat actor targets

North America was the most heavily targeted region in 2025, appearing in 82.04% of observed incidents. EMEA was a distant second place with 13.97%, and APAC came third with 3.99%.

The most targeted industries in incident response engagements were manufacturing, business services, and retail. This follows trends seen elsewhere, with both business services and manufacturing featuring heavily in ransomware leak posts across 2025.

Manufacturing was most heavily targeted by malware, social engineering, and account compromise/Business Email Compromise (BEC). The top forms of malware include Bunny Loader, ClickFix, and trojanized/renamed tools. VPN with no MFA was observed across multiple incident response investigations.

Business services attack vectors for 2025 mirror manufacturing, with malware, social engineering, and account compromise/BEC being the most commonly seen. Bunny Loader, renamed remote access tools, SocGholish, and ClickFix were the top files observed.

Social engineering, malware, and BEC lead the way for retail. In terms of top malware, Bunny Loader is, of course, present, with NetSupportRAT, Raspberry Robin, and Mintstealer making up some of the other top malware threats. VPN with no MFA was, as with manufacturing, observed in a majority of investigated incidents. Inconsistent MFA enforcement, and compromise via legitimate credentials, along with eventual deployment of several ransomware strains, were all features of 2025 incident response.

Selling Access in the Underground Marketplace

Now we pivot to Initial Access Brokers (IABs): specialized cybercriminals who compromise corporate networks to establish an initial foothold (e.g., a VPN account, RDP session, or web shell). IABs' core business is the sale of this unauthorized access to other criminal entities, most notably affiliates of Ransomware-as-a-Service (RaaS) operations.

This specialization enables sophisticated ransomware groups to outsource the complex network intrusion phase, allowing them to focus their resources exclusively on extortion.

IABs, often referred to as the "digital real estate agents" of the dark web, undertake the labor-intensive task of exploiting vulnerabilities and then transacting verified access credentials on underground forums.

To get a better understanding of their patterns of exploitation, we browsed threads from the past six months published on the most notorious cybercrime forums: DarkForums, Breached, XSS, Exploit[.]in, and RAMP.

The most active forum in the past six months was DarkForums with 221 access offerings, followed by RAMP with 208 (Figure 8).

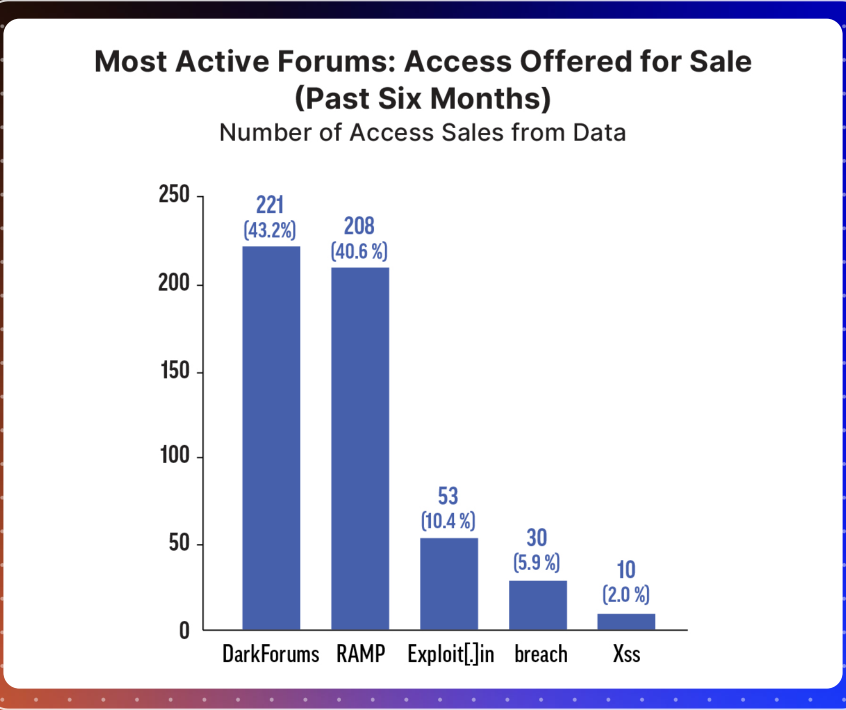


Figure 8

The average alleged revenue of the organizations whose access is being sold in these forums topped $3.2 billion USD, and the average base price was $113,275.

Remote Desktop Protocol (RDP) was the most frequently advertised access type, representing 21.2% of forum listings, followed by Virtual Private Network (VPN) at 12.8%, and RDWeb at 11.2%. These findings are consistent with our incident response observations, indicating that much of the access threat actors are achieving is likely ending up for sale in an IAB forum.

Privileged access adds value to the IAB's sale. When we analyzed the various privilege types offered, we found that the "domain user" and "domain admin" privilege levels were the most commonly offered, together accounting for roughly 75% of the privileges offered (Figure 9).

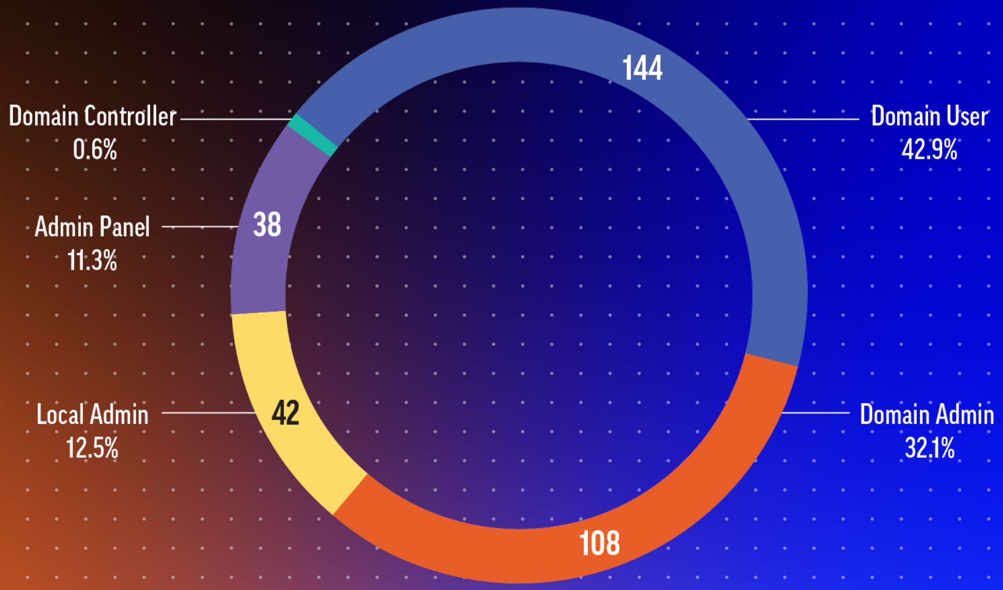


Figure 9

DarkForums and RAMP were identified as the most concentrated source of activity for the exchange of network access. Geographically, the United States remains the primary target, accounting for 30.9% of global illicit network access listings.

IAB activity is primarily concentrated on sectors offering the highest potential for financial gain or intelligence acquisition. The most targeted industry verticals are government (14.2%), retail (13.1%), and Information Technology (10.8%). Where the government sector is concerned, "admin panel" access is the most commonly observed type offered, with DarkForums serving as the principal platform for this type of sale. Retail is an attractive target due to a combination of payment card information (PCI) and personally identifiable information (PII), and the IT sector is similarly valuable to threat actors given its potential as a supply chain vector for a wide range of targets.

EXPOSURE SURFACES AS A STRATEGIC TERRAIN

The sophistication of 2025's threats lies in their architecture. Adversaries are building resilience through decentralized infrastructure, polymorphism via AI, and the co-opting of trusted platforms.

State-nexus actors conceal themselves with ORB networks

The shift from traditional botnets to Operational Relay Box (ORB) networks represents a sophisticated evolution in cyber espionage. Throughout 2025, we observed a surge in ORB adoption, particularly by state-nexus actors who require long-term, non-attributable access to sensitive targets.

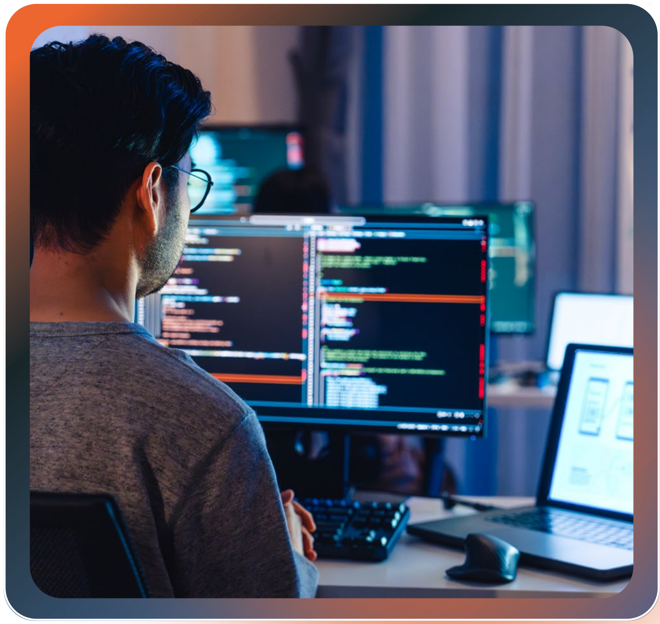
ORB networks are made up of compromised edge devices, such as Small Office/Home Office (SOHO) routers, firewalls, and IoT hardware, and are used as a private proxy mesh. While old-school botnets prioritize volume and "noise" to overwhelm targets, ORBs prioritize stealth, stability, and the obfuscation of state-sponsored activity.

Furthermore, unlike a standard botnet used for DDoS attacks or spam, an ORB network is designed to act as a chameleon layer. It sits between the threat actor and their victim, making the malicious traffic look like legitimate, everyday internet traffic originating from residential or small business IP addresses.

For example, the LapDogs ORB network is not a massive, noisy botnet. It consists of approximately 1,000 carefully selected high-bandwidth nodes, primarily compromised SOHO routers and Linux-based edge devices from vendors like Ruckus, ASUS, and Mikrotik. These devices are used to proxy command traffic, creating a non-attributable mesh network that routes traffic between the operator and the victim.

This "just-In-time" code generation creates a nightmare for traditional security operations:

- **Bypassing hash-based detection** — Since the payload is generated in real-time and exists only in memory, there is no static file (hash) for EDR tools to blacklist.
- **Defeating pattern matching** — Standard behavioral strings (like Get-WmiObject) are absent. Because the AI varies its "style" for every request, signature-based detection becomes useless.
- **Creating a new detection frontier** — Defenders can no longer look for "what the malware is." They must now look for "what the malware is doing"; specifically, identifying the anomaly of an unauthorized process querying an AI API followed immediately by dynamic script execution.



Cloud hijacking places a trojan in the meeting

The APT group Earth Kurma has pioneered a "Living Off the App" strategy by weaponizing Cisco Webex as a covert Command-and-Control (C2) infrastructure. Their specialized toolkit, consisting of DOWNBEGIN and SIMPOWEBEXSPY, treats Webex "Rooms" as virtual staging areas for espionage. DOWNBEGIN polls these rooms via the legitimate Webex API to receive encoded instructions, while SIMPOWEBEXSPY automates the theft of sensitive documents by uploading them as standard file attachments directly into the collaboration platform.

This technique is exceptionally difficult to detect because the malicious activity is indistinguishable from routine business operations. The traffic is encrypted, authenticated with valid API tokens, and destined for trusted Cisco domains, allowing it to bypass standard firewalls and Data Loss Prevention (DLP) sensors. By hiding within the noise of daily enterprise communication, Earth Kurma ensures that their data exfiltration and command traffic appear as nothing more than legitimate employee collaboration.

Modern API and supply chain infrastructure weaponized

Adversaries are increasingly exploiting GraphQL APIs, leveraging their inherent complexity and introspection features to mask malicious intent. A major 2025 supply chain attack targeted this ecosystem via the "GraphQL Network Inspector" Chrome extension. After hijacking the developer's account, threat actors pushed a malicious update (version 2.22.6) containing obfuscated scripts. These scripts silently harvested session cookies, API keys, and sensitive authentication tokens directly from the browsers of developers using the tool to debug their own production environments.

Beyond credential theft, attackers are also weaponizing the structural logic of these APIs to cripple infrastructure. A notable exploit in GitLab (CVE-2025-12562) demonstrated how unauthenticated users could bypass query complexity limits within the GraphQL API. By submitting deeply nested or resource-intensive queries that the system failed to properly throttle, attackers could trigger massive resource exhaustion. This effectively turned a sophisticated query language into a tool for highly efficient Denial of Service (DoS) attacks against critical DevOps infrastructure.

Dedicated malware for ICS environments

In 2025, the emergence of FrostyGoop malware marked a chilling milestone in industrial control systems (ICS) warfare, as it became the first malware specifically tailored to disrupt physical infrastructure by manipulating the Modbus TCP protocol. This malware was deployed in devastating cyber-physical attacks against district heating companies in Ukraine, cutting off essential heating services for civilians during sub-zero temperatures. Unlike traditional exploits that target software bugs, FrostyGoop targets the industrial process itself, sending legitimate Modbus commands to read and write to holding registers on controllers. By altering these operational parameters, attackers can force industrial systems into unsafe states or trigger complete shutdowns without ever tripping standard security alarms.

This evolution signifies a critical shift from Living Off the Land in IT environments to "Living Off the Protocol" in OT sectors. Because FrostyGoop uses authorized, standard commands, traditional security tools that only check for protocol compliance or known vulnerabilities are rendered ineffective. To counter this threat, defenders must move toward semantic monitoring, analyzing not just whether a command is "legal" according to the protocol, but whether its intent is safe for the physical process. This shift requires deep integration between cybersecurity monitoring and the engineering logic that governs industrial safety.

AI AS AN ACCELERATION LAYER

In 2025, generative AI shifted from “novel tooling” to a legitimate force-multiplier across the threat landscape. What we saw most consistently wasn’t brand-new “magic” attacks, it was adversaries bolting AI onto proven playbooks to move faster, scale wider, and reduce operator skill requirements. OpenAI’s threat reporting over 2025 characterized this pattern clearly: models were used to accelerate phishing content creation, scripting, and iterative problem-solving rather than unlock wholly new offensive capability. Others similarly documented broad experimentation by both cybercrime and government-backed actors with AI across the attack lifecycle (recon, initial access, development, and influence operations).

Monitoring outbound calls to AI APIs, protecting API tokens, and logging agent and tool actions became core detective controls in 2025, not “nice-to-haves.” As adversaries began experimenting with adaptive and AI-assisted malware, SOCs increasingly faced volumes and speeds of activity that exceeded human-only analysis. This shift reinforced the need for defenders to embrace the same AI technologies, using machine learning and generative AI to triage alerts, correlate telemetry across domains, summarize investigations, and surface weak signals at machine speed.

Effective defense in 2026 requires pairing traditional detection engineering with AI-augmented SOC workflows, ensuring defenders can match adversary acceleration rather than fall behind it.

The AI attack surface expands faster than controls

The most dangerous aspect of AI in security isn’t only what attackers are doing today, but that the AI attack surface is exploding (agents, tool use, connectors, plugins, model servers, prompt pipelines) while defenders struggle to inventory, test, and constrain it at the same speed. In practice, this leads to the failure of classic security controls. Excessive privilege, exposed services, unsafe deserialization, weak secret handling, all now appear inside LLM/agent stacks, often deployed quickly and operated by teams without mature security ownership.

Here is where we’re currently seeing AI leveraged in real intrusions:

- **Social engineering at industrial scale** — As noted in Figures 6 and 7, our observations show that AI has materially improved speed and personalization of lures, especially multilingual phishing, pretexting, and “brand-accurate” impersonation. This aligns with public reporting that threat actors increasingly use automation and AI tooling to make campaigns more convincing and to iterate quickly. In ransomware ecosystems, multiple observers described automation (including AI-assisted workflows) shrinking the time from initial access to broader compromise, aka compressing the defender’s response window.

- **Ransomware operations** — AI is increasingly being embedded into ransomware workflows to refine targeting, accelerate analysis of stolen data, and scale extortion operations. For example, the emergence of LAMEHUG, a toolset utilized by the threat actor Pawn Storm, marks a pivot toward "generative espionage." By leveraging a "Bring-Your-Own-AI" (BYOAI) model, the malware moves away from static, predictable payloads toward polymorphic, AI-generated code that adapts in real-time.

- **“Shadow AI” and data leakage: prompts became an asset** — In 2025, “prompt/data exhaust” became a new collection surface as organizations adopted dozens of genAI tools. We saw increasing discussion of attackers targeting AI-related data stores and user interactions (e.g., exfiltrating chatbot conversations via malicious browser extensions), because those chat logs can contain credentials, sensitive internal context, or step-by-step operational procedures.

AI infrastructure vulnerabilities actively exploited

While AI itself was rarely the direct exploit vector, AI infrastructure rapidly became a high-value target in 2025. Model servers, orchestration frameworks, and LLM web interfaces were often deployed quickly, exposed to the internet, and secured like internal tooling, creating familiar but dangerous failure modes.

The most impactful vulnerability patterns we have observed across AI stacks include:

- **Unsafe deserialization and memory handling in model servers** — Vulnerabilities in high-performance inference frameworks (e.g., vLLM and NVIDIA Triton) enabled denial-of-service conditions and, in some cases, risked remote code execution through crafted requests or model artifacts.

- **Weak authentication and token exposure in LLM platforms** — Self-hosted AI platforms and model runners (such as Ollama-based deployments) exposed API tokens or allowed authentication bypass, creating straightforward takeover paths when combined with default configurations.

- **Arbitrary file access in AI web interfaces** — Popular AI front-ends and demo frameworks exposed file-copy or file-read primitives, enabling attackers to stage sensitive data, crash services, or pivot to further compromise.

- **Serialization and injection flaws in agent and orchestration frameworks** — Frameworks designed to “chain” tools and actions (e.g., LangChain ecosystems) inherited classic injection risks, made more severe by the fact that LLM-generated content frequently crosses trust boundaries inside applications.

- **Supply chain compromise: The "Postmark-MCP" incident** — Attackers have begun poisoning the MCP ecosystem. A malicious version of the postmark-mcp package (version 1.0.16 and higher) was discovered on npm. This package, intended to let AI agents send emails via Postmark, contained a backdoor that secretly BCC'd every email sent by the AI agent to an attacker-controlled domain. This represents a "rug pull" attack where a trusted tool is weaponized post-deployment.

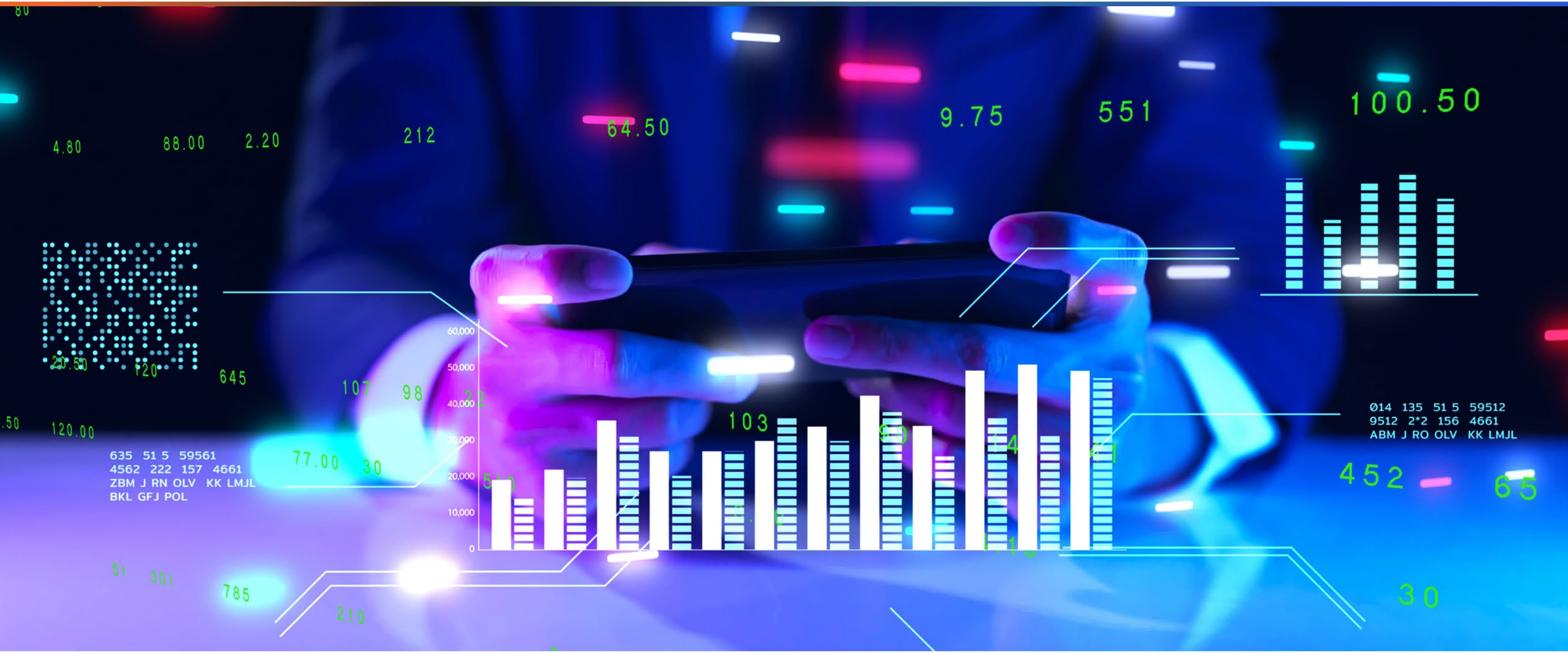
All of the above were not exotic bugs. They were well-understood vulnerability classes appearing in a new, fast-moving technology stack that attackers increasingly view as soft, high-impact targets. As AI systems become embedded into security operations, business workflows, and data processing pipelines, flaws in AI infrastructure increasingly represent privileged execution paths, not experimental side projects.

RANSOMWARE AS A DOWNSTREAM INCOME

Access isn’t all that’s for sale on the dark web. In 2025, 42% of our MDR incident response investigations involved ransomware, making it the single most common operational outcome observed.

The market for personally identifiable information (PII) is foundational to identity theft, with "Fullz" being the primary commodity for fraudsters. These comprehensive identity packages contain a victim’s full name, address, Social Security Number (SSN), and date of birth, typically selling for between $20 and $100 based on the freshness of the data and the victim’s credit score. Individual identity components are sold at lower tiers, with standalone US SSNs priced from $1 to $6, while scanned identity documents such as driver's licenses and passports command between $70 and $165 depending on the country of origin.

Financial data, particularly credit and debit card dumps, remains highly sought after for direct monetization. A standard US credit card with its CVV typically sells for $10 to $40, though cards verified to have high credit limits (over $5,000) can fetch up to $120. Regional differences are significant; cards from the UK and Germany are often priced higher (up to $60) due to stricter fraud detection and rarer supply.



In early 2025, a massive breach involving 270,000 customer identities from a German electronics brand was sold for a total of $250, illustrating how data saturation from large-scale leaks can drive bulk prices down.

Specialized logins for online banking and cryptocurrency exchanges are valued for the immediate liquidity they provide. Access to a retail consumer's bank account is priced between $150 and $500, but credentials for accounts with verified balances exceeding $100,000 can command several thousand dollars. Similarly, verified crypto exchange accounts on platforms like Kraken or Coinbase sell for $120 to $1,170, with the highest prices paid for accounts that are fully "laundering ready" and have high-level Know Your Customer (KYC) verification.

Our observations indicate that the trade has increasingly shifted toward "infostealer logs," which harvest entire browser environments, including saved passwords and session cookies. These logs sell for an average of $10 each and are particularly dangerous because stolen cookies allow attackers to bypass multi-factor authentication (MFA) by hijacking active sessions. This market is now highly industrialized, with criminals purchasing "bulk cloud subscriptions" for $200 to $500 per month to receive a continuous stream of fresh logs delivered via private Telegram channels.

Leak site posts convert ransomware to income

2025 was a year of continued escalation and power-base solidification by major ransomware threat actors. One last burst of activity at the end of the year served to push leak post numbers higher than they'd been up to that point, edging out Q1's high of 1,611 with a Q4 close of 1,661. This more than made up for Q2's slight dip of 871, with Q3 bouncing back up to 943.

This increase in leak posts was accompanied by Ransomware-as-a-Service (RaaS) and double extortion cementing themselves as go-to tactics for threat actors large and small. The emergence of “collectives” (experienced threat actors joining forces and layering their expertise in initial access, ransomware deployment, and data exfiltration) has become a persistent cause for concern when planning out defensive strategies.

Key industries have remained under fire all year long, with threat actors particularly focused on healthcare, business services, and manufacturing where double extortion is concerned.

Qilin leak post total far surpasses other top ransomware groups

We're able to learn a lot by tracking the groups making the greatest number of leak site posts. The top 10 ransomware groups of 2025 (Figure 10) reflect the trends Rapid7 has observed over much of the previous 12 months.

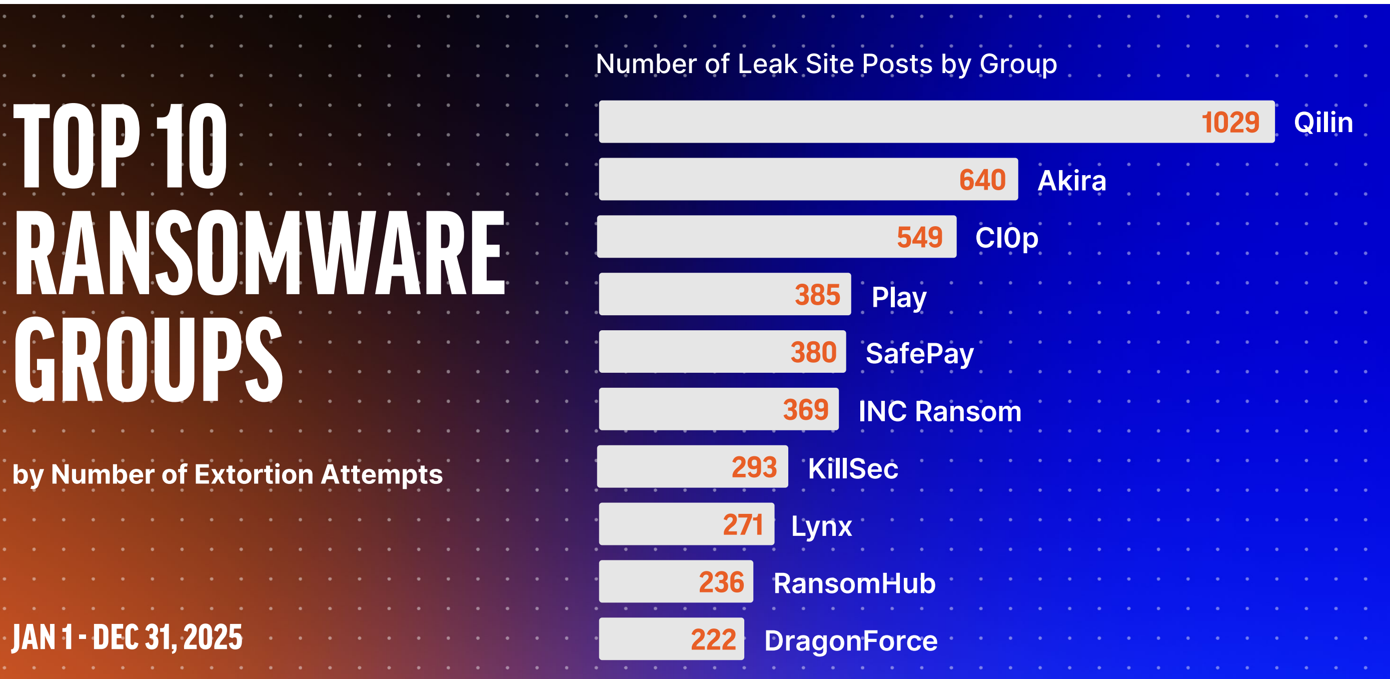


Figure 10

In February 2025, Qilin began moving up the ranks among our top 10 ransomware groups. It reached the number one spot in May and maintained that position, with the exception of a brief overtaking by Cl0p in November, for the remainder of the year. Qilin's double-extortion tactics, combined with a successful RaaS business model, are making it a significant threat. The cost to victims has ranged from exfiltration of large amounts of sensitive information, to disruption and shutdown of services in the most severe cases.

Overall, we saw an average of 56 active groups per month in 2025, with 140 unique active groups in the entire dataset (versus 102 unique groups in 2024), and 78 new groups not seen previously. The total number of posts overall rose from 6,034 in 2024 to 8,835 in 2025, a rise of 46.4%. The number of unique ransomware groups grew from 102 in 2024 to 140 in 2025, with average posts per group also increasing slightly: 59.2 posts per group in 2024, versus 63.1 posts per group in 2025.

This 2025 data highlights the continued expansion of the ransomware ecosystem, thanks to numerous threat actors entering the space alongside a notable rise in leak post output. There is no slowing down here; rather, there is a sustained pace of operations and a dynamic, always-shifting edge to the groups themselves. With so many threat actor specializations to contend with, smart use of accurate ransomware threat intelligence has never been more important.

Business services, manufacturing, and healthcare a prime target

Business services, manufacturing, and healthcare retain the same top three places from Q3, broadening these positions out into the top three targeted industries across 2025. Construction and technology are in fourth and fifth place, with legal, finance, and retail taking the next three places. Hospitality and education round out the most popular industry targets of 2025.

The top regional targets of 2025 solidify what Rapid7 analysis has evidenced all year long. The US is, by far, the most heavily targeted region of 2025 with 70% of all observed leak posts. Second place goes to Canada (6%), with the UK and Germany both sitting at 5%.

Threat actors know they stand the best chance of big payouts by targeting US sectors which are wealthy, provide critical services, or take up important pieces of supply chains. A supply chain attack specifically could ripple out across countless other industries and services. Sectors that are filled with legacy or bespoke systems that by default aren't easily upgradeable due to cost or complexity, will also be appealing targets.

The healthcare, education, manufacturing, and business services sectors contain significant amounts of sensitive data, financial information, and business critical data, all of which command high prices during double-extortion negotiations. Many of the top threat actors listed have displayed a fondness for these sectors for several years, and this is unlikely to change.

The RaaS ecosystem matures and industrializes

A top-level trend Rapid7 has observed over time has been the continued industrialization and maturation of the cybercrime ecosystem. Nowhere has this progression been more pronounced than in ransomware-as-a-service (RaaS). This trend is characterized by a professionalized supply chain, increasing tactical sophistication, and sharpened focus on high-value exploits.

Groups shift from slick branding to tactical evasion

While law enforcement continued to raise the temperature on ransomware groups, some threat actors attempted to dial down the heat. In H1, Rapid7 Labs observed that new groups were using glossy branding and splashy attacks to gain visibility, and perhaps be snapped up by major RaaS operations. Anubis, with slick branding and social media hyping of attacks, stood out in this regard. H2 saw a reversal of this trend, with newer and smaller groups removing URLs and other identifiers from their data leak teasing.

Perhaps these groups wanted to evade law enforcement action, or simply avoid being infiltrated and detailed by security researchers. Whatever the reason, some groups chose to deliberately end the year with a visibility whimper, instead of a bang.

RaaS and double-extortion tactics maximize gains

We observed many ransomware groups leaning heavily on favored tactics in 2025. Some allowed their affiliate networks to do the talking, while others ventured into the cloud, or advertised their wares on auction sites to maximize ill-gotten gains.

RaaS and double extortion were firm favorites, featuring heavily in every quarter's Top 10. Almost every threat actor in 2025's overall Top 10 makes use of one or both, with few exceptions. The primary standout here is SafePay, a group originating in 2024 which avoids RaaS entirely in favor of running its own self-contained operation.

While RaaS-free and fileless ransomware groups will continue to gather victims, RaaS is currently the dominant force among ransomware threat actors. This will almost certainly be the case for some time to come, as threat actors invested in this approach continue to refine their tactics and expand their operations.

Affiliate drift and collaboration make for uneasy extortion alliances

Affiliate drift and alliances which seemed as though they may break at any moment were observed throughout 2025, likely spurred on by restructuring activities. Many affiliates moved to RaaS offerings, encouraged by features such as Qilin's "call a lawyer" service, or Lynx's slick affiliate sections containing victim profiles, executable archives, and news pages.

Threat actors which largely operated alone, or were not particularly known for collaboration, started to pool resources and work together. This is perhaps best illustrated by the formation of "Scattered LAPSUS$ Hunters," a collective made up of Scattered Spider, ShinyHunters, and LAPSUS$. Each of these groups offered up unique skills and specializations, from phishing to data exfiltration, and their group — sometimes referred to as an "extortion alliance" — was responsible for many major breaches worldwide.

Some groups, most notably DragonForce and Cicada3301, reduced fees or more general barriers to entry, in an effort to attract new affiliates. It's not unheard of for affiliates to go rogue, generating unwanted attention for RaaS operations either through visible activity or leaking internal secrets. This brittle environment, made more confusing by infighting and power struggles (DragonForce's "hostile takeover" of RansomHub in Q2, for example), made for a tense 2025.

High-profile exploits of the cloud, ESXi, and major enterprise suites

Elsewhere, threat actors made use of open source software such as TruffleHog, a tool designed to detect and highlight insecurely stored credentials in GitHub repositories, to hunt down cloud credentials. The Rapid7 Incident Response team observed two examples of Crimson Collective making use of TruffleHog in September, creating new users and escalating privileges once inside the network. ESXi, VMware's bare-metal hypervisor software, also became a popular target.

Cl0p made waves with a devastating coordinated campaign which targeted users of Oracle's E-Business suite. The latter involved potentially months of undetected access, and emails sent to targets threatening to leak stolen data. Curiously, the exploit likely used in these attacks may have been leaked in a Telegram channel accidentally.

We observed threat actors favoring initial access techniques leaning toward rapid data exfiltration and heightened pressure for the victims, as highlighted earlier in this report. From ransomware to incident response, software which enables file sharing, transfer, and collaboration are frequently key targets in the race to breach the network and extract business critical data.

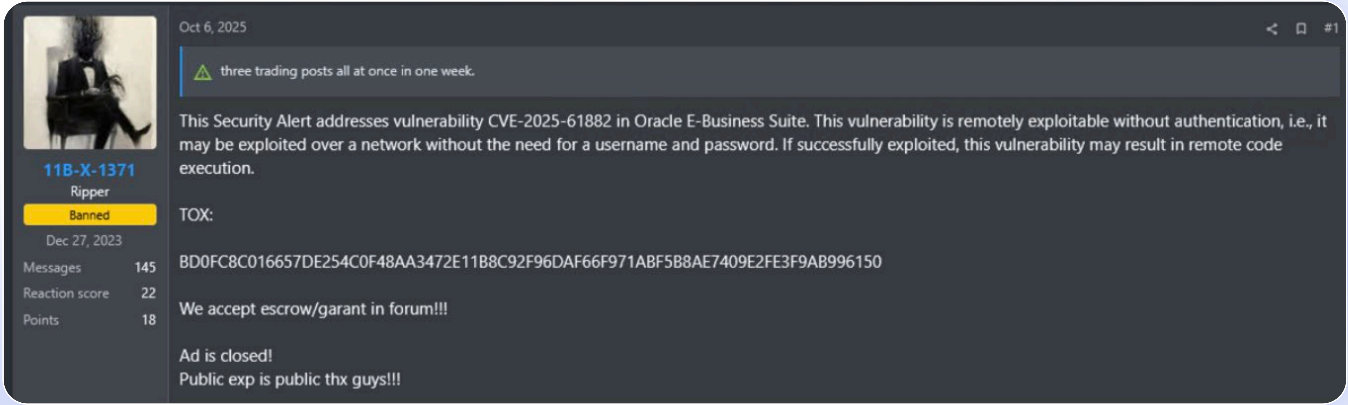
Data auctions have also become part and parcel of some threat actors' strategy, most notably Rhysida and Warlock making use of double extortion and the possibility of selling off stolen data if ransom demands are not met.

An underground market for tools of the trade

Threat actors generate additional income by offering their exploits, tools, services, and malware at a price on the underground market.

The 2025 ecosystem for cybercrime tools is defined by professionalized "Malware-as-a-Service" (MaaS) platforms that lower the technical barrier to entry. A prominent example is "Olymp Loader," an assembly-based malware that surfaced in mid-2025 and is marketed for its anti-analysis and anti-detection capabilities. These tools are sold via subscription models on forums like XSS and HackForums, providing attackers with modular features such as stager generators, botnet management, and built-in stealer modules for browser data and cryptocurrency wallets.

High-end exploit chains for zero-day vulnerabilities command the highest prices in the subterranean economy, often used by state-sponsored actors and sophisticated ransomware groups. Exploit brokers like Crowdfense currently pay between $5 million and $7 million for zero-click full-chain exploits for iPhone or Android devices. Remote code execution (RCE) vulnerabilities in desktop browsers like Chrome and Safari are also highly valued, with prices ranging from $2 million to $3.5 million. An exploit targeting a vulnerability in the Oracle E-Business Suite (CVE-2025-61882) was observed being offered for sale in the RAMP forum (Figure 11).



11B-X-1371
Ripper
Banned
Dec 27, 2023

Messages	145
Reaction score	22
Points	18

Oct 6, 2025 #1

⚠ three trading posts all at once in one week.

This Security Alert addresses vulnerability CVE-2025-61882 in Oracle E-Business Suite. This vulnerability is remotely exploitable without authentication, i.e., it may be exploited over a network without the need for a username and password. If successfully exploited, this vulnerability may result in remote code execution.

TOX:

BD0FC8C016657DE254C0F48AA3472E11B8C92F96DAF66F971ABF5B8AE7409E2FE3F9AB996150

We accept escrow/garant in forum!!!

Ad is closed!
Public exp is public thx guys!!!

Figure 11

CVE-2025-61882 represents a critical vulnerability within the Oracle E-Business Suite (versions 12.2.3–12.2.14). This flaw enables unauthenticated attackers to execute arbitrary code through HTTP, thereby enabling complete system compromise. The vulnerability has been exploited as a zero-day by Cl0p to exfiltrate financial and human resources data for subsequent extortion attempts.

To evade modern security solutions, attackers utilize fully undetectable (FUD) “crypters” (aka encrypters) which obfuscate malicious code to bypass EDR and antivirus detection. These crypters typically cost around $100 for a monthly subscription, though specialized versions like "Nightmangle" offer lifetime access for $999 and include a 24-hour trial period. Some crypters are specifically advertised for their ability to bypass major defenses like Windows Defender, Kaspersky, and ESET.

Resilient hosting is provided by bulletproof hosting (BPH) providers who ignore abuse complaints and legal requests. These providers often utilize advanced techniques like fast-flux BGP and ASN hijacking to maintain uptime, with servers located in nuclear-protected bunkers in jurisdictions like Russia and the Netherlands. Bulletproof VPS hosting is sold at a premium, typically costing $20 to $50 per month, which is significantly higher than the $2 to $10 charged for legitimate commodity hosting services.

How to stay out of ransomware’s cross-hairs in 2026

Tactics used by ransomware threat actors in 2025 did not change significantly in terms of innovation or a unique evolution of techniques. If anything, the favoritism shown toward certain intrusion methods indicates key areas where defenders can focus their attention in 2026.

- Social engineering is a popular choice for major ransomware threat actors, as well as being a key tactic deployed against the top targeted sectors in Rapid7’s incident response data. Locking down the help desk, as well as limiting high-risk password resets, will both help here.

- Making strict use of correctly configured MFA controls for critical systems, remote access points, and privileged accounts will (for example) prevent attackers from gaining easy access via insecure RDP and VPN. Limit push attempts, and enable number matching, to ward off MFA fatigue attacks.

- Spam filtering will help to reduce ransomware attacks pinned to social engineering, and user awareness training will educate users about the risks of convincingly crafted emails bearing malicious attachments. Training will also help in the fight against Business Email Compromise (BEC), often deployed against specific industries as observed in Rapid7 Incident Response investigations.

- Network edge devices are a favored method of initial access for many types of threat actors, and ransomware groups are no exception. Continuous patch management, prioritizing fixes by known exploits, and potential risk to your organization, is key.

- While data may be exfiltrated despite your best efforts, knowing business operations won’t crash to a halt thanks to backups — and knowing you won’t need to pay a ransom — will help you through the early stages of a confirmed breach. Implementing immutable backups, regularly testing recovery procedures, and ensuring backups are isolated from the production network, will make all the difference. So too will being able to rapidly invalidate active sessions and tokens, and forcing enterprise-wide password resets alongside maintaining control of help desk password resets.

EMBEDDED ACCESS AND PRE-POSITIONING CREATES GEOPOLITICAL WAVES

The cyber threat landscape of 2025 marks a definitive departure from the paradigms of previous decades. We are no longer observing a distinct separation between state-sponsored espionage and destructive cyber warfare, nor are we seeing a clear delineation between advanced persistent threats (APTs) and the tools of cybercrime.

Instead, 2025 has been defined by the convergence of these domains into a unified theater of hybrid operations.

Nation-state activities highlight a shift to pre-positioning of assets

Nation-state actors, particularly those aligned with the People's Republic of China (PRC) and the Russian Federation, have shifted their operational focus from pure intelligence gathering to the strategic pre-positioning of assets within critical infrastructure for the kinetic effects. Simultaneously, the democratization of advanced offensive capabilities — fueled by the weaponization of artificial intelligence (AI) and the proliferation of "One-Day" exploits — has empowered cybercriminal syndicates to execute intrusions with state-like sophistication.

China is guided by a doctrine of pre-positioning, which involves embedding persistent access into critical infrastructure for intelligence collection today and optional disruption during future geopolitical crises. The Volt Typhoon group exemplifies this by using extreme operational security and Living Off the Land techniques to maintain stealthy persistence within critical sectors like energy and telecommunications, aiming for covert footholds rather than immediate sabotage.

Further illustrating China's focus on deep access, Salt Typhoon targets the global telecommunications backbone, including lawful interception systems, to conduct large-scale traffic collection without deploying noisy endpoint malware. Other groups like Earth Kurma and Superjumpers blend espionage with legitimate cloud service abuse, leveraging platforms like Cisco Webex and Operational Relay Box (ORB) infrastructure to create encrypted command-and-control (C2) channels embedded within normal enterprise workflows.

Russian cyber operations integrate sustained intelligence collection with the preservation of disruptive capabilities. Intelligence-focused groups like APT29 refine cloud-centric intrusion

techniques, utilizing credential theft, access token abuse, and manipulation of trusted SaaS and identity relationships to achieve extended dwell times with minimal forensic footprint. This includes campaigns that leverage Microsoft 365 device-code authentication abuse to capture access without traditional password-based intrusion paths.

The evolution of Russian tradecraft is seen in Pawn Storm (APT28), which now uses LAMEHUG malware to dynamically generate command-line instructions via Large Language Model (LLM) APIs, effectively reducing static signatures. Meanwhile, military-aligned actors like Sandworm focus on strategic sabotage, deploying destructive tools like PathWiper to target critical infrastructure, reinforcing the concept of operational readiness as a key component of their integrated power model.

Iranian nexus threat actors show increased operational adaptability, with a recurring emphasis on exploiting internet-facing infrastructure, network-edge devices, and cloud-hosted services. These paths provide access in environments with reduced defensive visibility, supporting extended dwell times. Their campaigns prioritize persistent access, intelligence collection, and strategic positioning through the rapid rotation of C2 and the use of compromised third-party hosting.

North Korean (DPRK) activities are structurally distinct, driven primarily by financial generation and sanctions evasion alongside intelligence objectives. Their defining feature is a heavy reliance on people-centric access — such as developer-focused social engineering, recruitment lures, and fake job offers — to compromise individuals. This approach enables direct financial theft, as seen in the $1.5 billion Bybit theft, and also introduces downstream supply-chain risk via trojanized coding assessments.

Finally, DPRK groups like Earth Kumiho (Kimsuky) emphasize "Weapons on the Cloud," hosting encrypted payloads in public GitHub repositories and using legitimate SaaS platforms such as Dropbox for C2 and exfiltration. This tactic embeds their malicious traffic within normal, encrypted enterprise cloud usage, effectively rendering traditional network-based blocking largely ineffective for sustaining their espionage and financial extraction at scale.

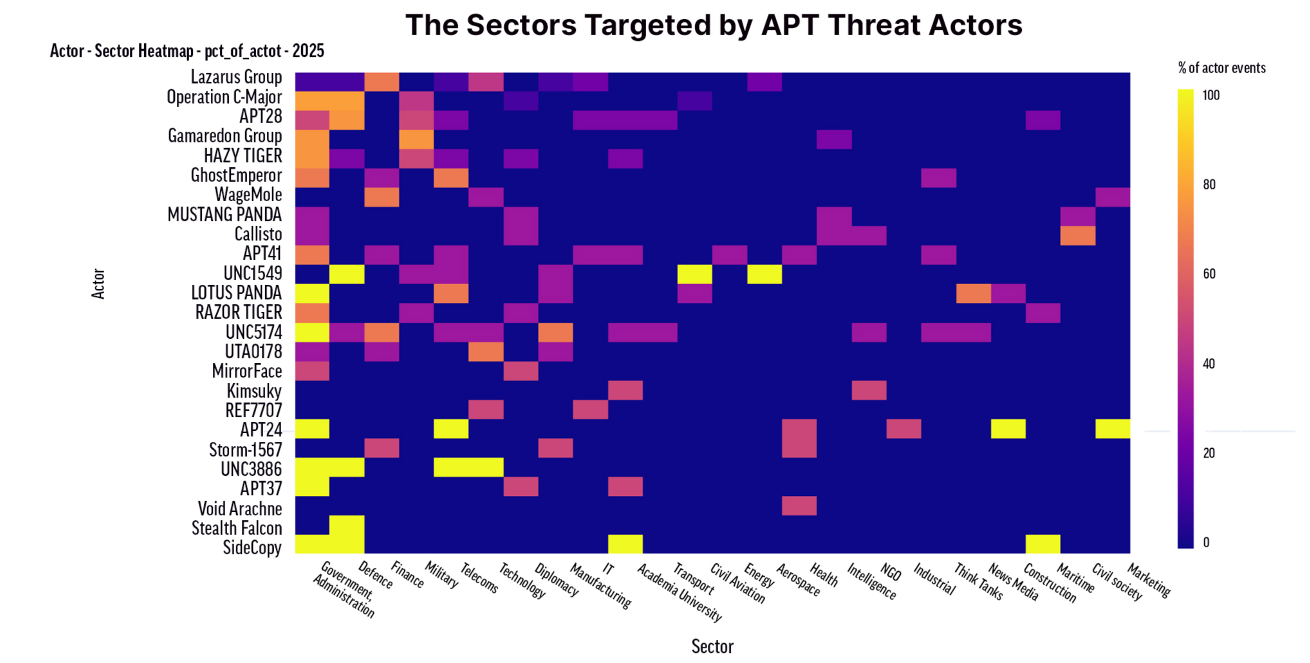


Figure 12

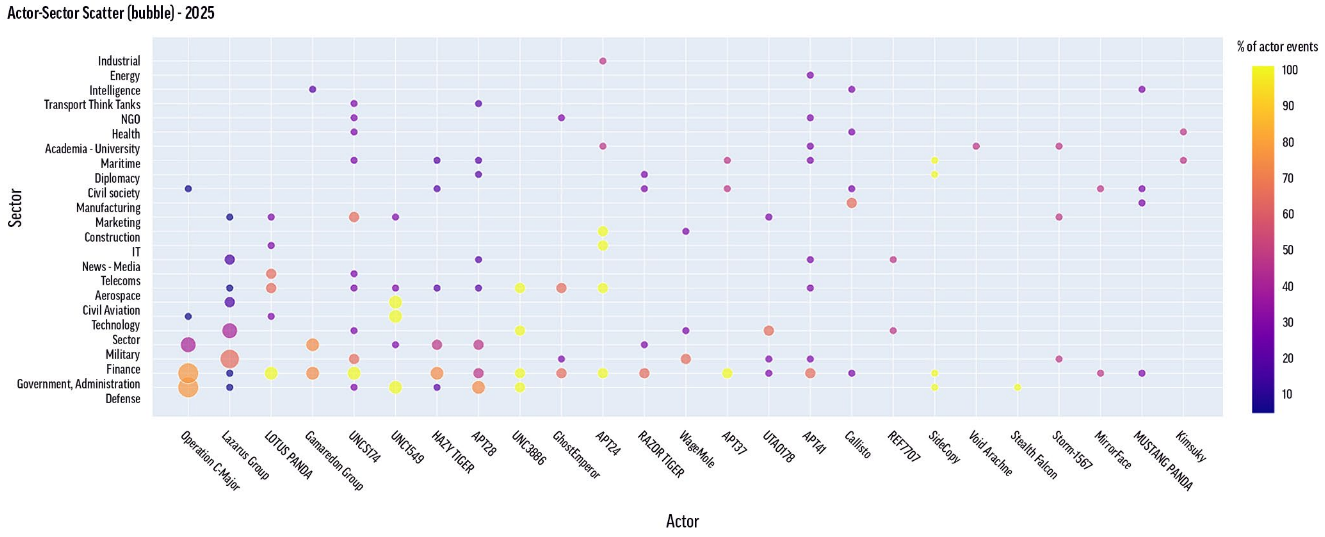


Figure 12

Hacktivists and politically-motivated actors are professionally crowdsourced

Alongside state-aligned operations, 2025 featured sustained and highly visible activity from geopolitically aligned hacktivist ecosystems, particularly during periods of heightened international tension. Pro-Russian denial-of-service collectives, notably NoName057 (16), conducted repeated and coordinated campaigns targeting government portals, public-sector services, election-related infrastructure, transportation-related systems, and media organizations across multiple European countries. Activity patterns frequently aligned with geopolitical flashpoints, including elections, sanctions announcements, and security-related incidents.

While these operations historically relied on volumetric DDoS tactics, late 2025 saw a shift toward more sophisticated application-layer (Layer 7) attacks. By targeting specific functions of web applications rather than just flooding bandwidth, these groups successfully bypassed traditional Content Delivery Network (CDN) protections. This evolution, fueled by the DDoSia project, signals a transition from a "limited technical sophistication" model to a "crowdsourced professional" model. Despite the lack of deep lateral movement or persistence, these campaigns generated a measurable operational impact, including service outages, public-facing disruptions, and a sustained incident-response burden for affected organizations.

An emerging arena of strategic competition

Looking toward 2026 and beyond, geopolitical cyber risk will be shaped less by isolated crises and more by a world entering a prolonged period of instability. Multiple regions are already volatile, and current and emerging conflicts will increasingly spill into the digital domain, directly affecting global digital infrastructures. Cyberspace is no longer a secondary theater but a permanent arena of strategic competition, where state-aligned actors seek long-term advantage by degrading resilience, shaping influence, and positioning themselves for future escalation.

As instability deepens, cyber operations will grow more virulent and systemic in nature. Rather than focusing on individual organizations, adversaries will exploit global interdependencies, fragile supply chains, shared platforms, identity systems, and edge infrastructure to achieve persistent access while minimizing attribution and escalation risks. Disrupted trade routes, contested maritime corridors, and energy chokepoints will further elevate logistics, transportation, and industrial ecosystems as high-value cyber targets used to apply economic and political pressure rather than immediate disruption alone.

At the same time, the rapid deployment of autonomous and semi-autonomous AI systems will expand the attack surface through poorly governed machine identities and opaque processes. These weaknesses will increasingly be leveraged as strategic enablers within broader geopolitical campaigns. In response, organizations and governments alike will continue to prioritize geopolitically driven cyber risk, accelerating moves toward identity-centric security, resilience-focused architectures, and, in some regions, tighter sovereign control over critical digital infrastructure.

THE PATH FORWARD: FROM REACTIVE DEFENSE TO EXPOSURE MANAGEMENT

The 2026 cyber threat landscape is all about acceleration, fundamentally changing the defender's operational calculus. The balance of power has shifted from one defined by speed of response to one defined by speed of exploitation. Our findings confirm that the statistical buffer between vulnerability disclosure and confirmed exploitation (i.e., the "predictive window") has materially collapsed. High-impact vulnerabilities are now operationalized almost immediately, a trend evidenced by the 105% increase in confirmed CVSS 7–10 exploitation year over year. Risk is no longer an accumulation of latent debt; for the most critical flaws, it is an immediate, realized event. This seismic shift highlights a growing mismatch between the velocity of modern threats and defensive processes designed for slower, more predictable cycles.

This acceleration is enabled by two primary forces: the industrialization of the cybercrime ecosystem and the pervasive adoption of AI. Initial access brokers have commoditized pre-authentication access, allowing ransomware collectives to bypass the complex intrusion phase and focus on smash-and-grab data monetization. Simultaneously, adversaries have strategically pivoted to high-value, exposed surfaces such as identity systems, cloud control planes, and collaboration platforms, blurring the line between legitimate and malicious activity. AI is bolting speed and scale onto these proven playbooks, accelerating social engineering, shrinking the dwell time of ransomware operations, and dynamically expanding the attack surface within fast-moving AI infrastructure itself.

To effectively manage cyber risk in 2026, organizations must adopt a fundamental mindshift toward preemptive security. This means moving beyond a reactive, volume-based vulnerability management approach and embracing an exposure management model focused on informed prioritization and anticipation.

By reducing the known, preventable conditions attackers monetize before exploitation occurs, defenders can regain a measure of control. The data unequivocally proves that delayed response and misinformed prioritization are no longer merely costly; they are increasingly determinative of a breach. Success will be defined by the capacity to connect technical exposure to business impact and apply AI-augmented workflows to match the adversary's machine speed.

ABOUT RAPID7

Rapid7, Inc. (NASDAQ: RPD) is a global leader in AI-powered managed cybersecurity operations, trusted to advance organizations' cyber resilience. Open and extensible, the Rapid7 Command Platform integrates security data, enriching it with AI, threat intelligence, and 25 years of expertise and innovation to reduce risk and disrupt attackers. As a recognized leader in preemptive managed detection and response (MDR), Rapid7 unifies exposure and detection to transform the cybersecurity operations of more than 11,500 customers worldwide. For more information, visit our website, check out our blog, or follow us on LinkedIn or X.

RAPID7

SECURE YOUR

Cloud | Applications | Infrastructure | Network | Data

ACCELERATE WITH

Command Platform | Exposure Management | Attack Surface Management | Vulnerability Management | Cloud-Native Application Protection | Application Security | Next-Gen SIEM | Threat Intelligence | MDR Services | Incident Response Services | MVM Services

SECURITY BUILT TO OUTPACE ATTACKERS

Try our security platform risk-free - start your trial at **rapid7.com**