2026 State of Malware

The dawn of machine-scale cybercrime

THREATDOWN™

Strimm@



Table of contents

03 **Foreword**

04 **Executive summary**

05 **Threat landscape**

06 The cost of a cyberattack

08 Anatomy of an attack

17 Five key operating patterns

19 Ransomware

25 **How AI is reshaping cybercrime**

29 **Conclusion**

30 **How to protect your company**

Foreword

Automated attacks using malware are no longer the driving force in cybercrime. In 2025, attacks on businesses were dominated by hands-on-keyboard intrusions where human hackers used legitimate tools to infiltrate networks and compromise data. These attacks were smarter, more deliberate, and far stealthier than the mass infections of the past.

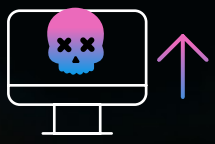
In 2025, the cost of cybercrime was no longer counted only in downtime or ransom payments, but in project delays, lost contracts, client churn, rising premiums, and punishing recovery costs.

Cyberattacks also exacted a broader toll on society, shutting down schools, steel mills, hospitals, and auto factories. But while organizations scrambled to adapt to the real-world havoc of human hackers, cybercrime had already embarked on its post-human future. In 2025, AI-driven ransomware emerged, AI-generated phishing and deepfakes became standard social engineering tools, and OpenAI CEO Sam Altman warned that AI had already "fully defeated" the biometric authentication used by banks.[1]

This report captures a unique moment of transition—when the established world of human-driven intrusion meets the emerging machine-driven future. It explores how the threat landscape is being reshaped by AI, what it means for defenders, and how organizations can prepare for a coming year in which human hackers' playbooks give way to tireless machine adversaries that learn, adapt, and scale on their own.

[1] C-SPAN (2025), OpenAI CEO Sam Altman Speaks at Federal Reserve Conference, https://www.c-span.org/program/public-affairs-event/openai-ceo-sam-altman-speaks-at-federal-reserve-conference/662859

Executive summary



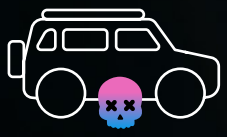
8%
increase
in ransomware attacks year-over-year



86%
of ransomware attacks
are conducted remotely



135
countries
experienced ransomware attacks in 2025



$2.5B
estimated
cost of the Jaguar Land Rover attack

The threat landscape is shifting rapidly from human-driven intrusions to AI-orchestrated attacks. In 2025, attacks on businesses were dominated by stealthy, hands-on-keyboard operations that relied on legitimate tools, stolen credentials, and human decision making, with little or no traditional malware. Attackers blended into normal administrative activity, moved quickly—often at night or on weekends—and exploited blind spots such as unmanaged systems to deploy ransomware remotely. By prioritizing speed, stealth, and the disruption of security and recovery mechanisms, they left defenders with shrinking windows to detect, contain, and respond.

By the end of the year, speculation about AI in cybercrime gave way to reality. AI systems outperformed human vulnerability researchers, exploit pipelines compressed patch-to-exploit timelines to minutes, and the first confirmed cases emerged of AI agents executing complex, multi-stage compromises with minimal human oversight. Cybercrime has entered its machine-scale era. Organizations that invest in visibility, resilience, and automation now will be best positioned to keep pace with a threat landscape evolving faster than ever before.

Five operating patterns every business should know:

1. **Faster attacks**
   Compressed dwell times leave businesses little time to respond.
2. **Working at night**
   Threat actors time their operations for periods of low visibility and reduced staffing.
3. **Living off the land**
   Legitimate tools and credentials are used to blend into normal activity.
4. **Staging from blind spots**
   Unmanaged and unmonitored systems are used to stage remote attacks.
5. **Attacking security and backup software**
   Security controls and recovery mechanisms are deliberately targeted.

Threat landscape

- The cost of a cyberattack
- Anatomy of an attack
- Five key operating patterns
- Ransomware

> "I haven't seen a malware infection case for 5 months."
>
> **– ThreatDown Threat Research Analyst**

The cost of a cyberattack

Businesses today operate within tightly linked supply chains and depend on a wide array of SaaS platforms and cloud services. These dependencies improve adaptability and efficiency while expanding attack surfaces and enabling breaches to cascade across entire ecosystems.

In this environment, data has become an organization's most valuable commodity, and cybercriminals make money from holding that data for ransom or selling it. The average ransom payment in the first half of 2025 was $750,000 USD[2] while the total cost of a data breach was six times higher at $4.4 million USD.[3] The commoditization of ransomware and a growing library of playbooks and turnkey attack methods means that small and medium size enterprises (SMEs) must repel the same threat actors as larger organizations with fewer resources and disproportionately greater consequences. Hiscox data from 2025 shows that 80% of SMEs hit by ransomware paid a ransom, but only 60% recovered their data. The fallout extended well beyond recovery: one in three faced substantial fines and nearly 30% reported declines in sales, customer trust, or new-business opportunities.[4]

Whatever the size of the target, in 2025, the cost of cybercrime was measured in more than dollars, and felt far beyond the networks and computers where they occurred.

> The average ransom payment in the first half of 2025 was $750,000 while the total cost of a data breach was six times higher at $4.4 million.

Technology blackout across 14 hospitals

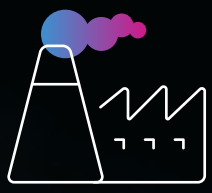


In May 2025, a ransomware attack triggered a technology blackout across Kettering Health's 14-hospital system in Ohio. Almost a terabyte of data was stolen, and thousands of doctors and nurses were forced to treat patients without medical histories, lab results, or medication records for nearly two weeks.[5]

[2] Calculation based on data based on Coveware (2025), Targeted social engineering is en vogue as ransom payment sizes increase, https://www.coveware.com/blog/2025/7/21/targeted-social-engineering-is-en-vogue-as-ransom-payment-sizes-increase; and Coveware (2025), Insider Threats Loom while Ransom Payment Rates Plummet, https://www.coveware.com/blog/2025/10/24/insider-threats-loom-while-ransom-payment-rates-plummet
[3] IBM (2025), Cost of a Data Breach Report 2025, https://www.ibm.com/reports/data-breach
[4] Hiscox (2025), Hiscox Cyber Readiness Report 2025, https://www.hiscoxgroup.com/hiscox-cyber-readiness-report-2025
[5] The HIPAA Journal (2025), Kettering Health Confirmed Patient Data Compromised in May 2025 Ransomware Attack, https://www.hipaajournal.com/kettering-health-ransomware-attack/

The cost of a cyberattack

20 steel mills shut down



Nucor, North America's largest steel manufacturer, was forced to temporarily shut down steel mills, recycling centers, and fabrication plants in the US, Canada, and Mexico in May 2025, after it detected unauthorized third-party access to some of its systems.

Grocery shelves emptied at 30,000 stores



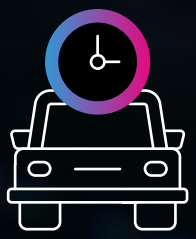
In June 2025, a cyberattack on United Natural Foods Inc. brought food supply to a standstill at 30,000 locations. Whole Foods stores, independent grocers, online retailers, and military commissaries faced empty shelves for weeks, resulting in an estimated $400 million USD in lost sales.[6]

Airline check-in attack leaves thousands stranded



In September 2025, a cyberattack on Collins Aerospace, a critical aviation technology provider, paralyzed check-in and boarding systems across major European airports including London Heathrow, Berlin Brandenburg, and Brussels, where 50% of flights were cancelled. Thousands of passengers were stranded, causing massive queues and delays throughout the weekend.[7]

Car production sent back to the 1950s

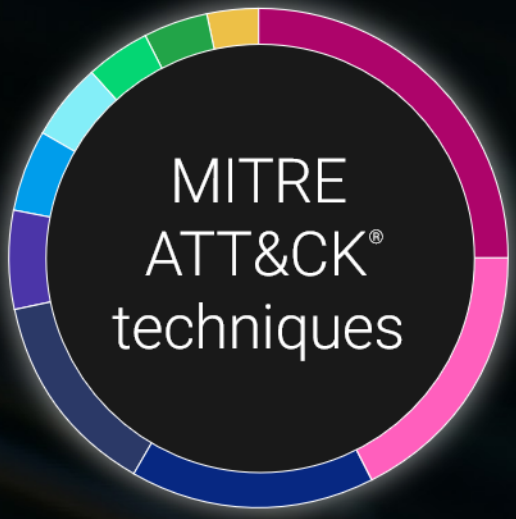


In August 2025, a cyberattack brought Jaguar Land Rover's (JLR) auto production in the UK, China, Slovakia, India, and Brazil to a halt for five weeks, costing JLR an estimated $50 million a week.[8] The attack affected 5,000 organizations across a complex supply chain, created a $2.5 billion hole in the UK economy,[9] and drove UK car production to its lowest level since the 1950s.

[6] Reuters (2025), United Natural Foods projects up to $400 million annual sales hit after cyber breach, https://www.reuters.com/markets/europe/united-natural-foods-projects-up-400-million-annual-sales-hit-after-cyber-breach-2025-07-16/
[7] CNN (2025), Disruption from cyberattack on European airports to stretch into Sunday, https://edition.cnn.com/2025/09/20/travel/europe-airports-cyberattack-intl
[8] BBC (2025), JLR cyberattack caused UK car production to hit 70-year low for September, https://www.bbc.co.uk/news/articles/cvgmp1prnv0o
[9] Cyber Monitoring Centre (2025), Cyber Monitoring Centre Statement on the Jaguar Land Rover Cyber Incident – October 2025, https://cybermonitoringcentre.com/2025/10/22/cyber-monitoring-centre-statement-on-the-jaguar-land-rovercyber-incident-october-2025/

Anatomy of an attack

In 2025, ThreatDown data showed adversaries increasingly relying on the same core tactics. Professional cybercriminals target entire organizations and attempt to ransom, leak, or sell their data. To avoid detection and maximize the chances of success, attacks are hands-on and executed in multiple phases, using little or no malware.



- 23.98% Defense Evasion – TA0005
- 16.81% Execution – TA0002
- 14.85% Discovery – TA0007
- 13.3% Persistence – TA0003
- 6.52% Initial Access – TA0001
- 5.67% Credential Access – TA0006
- 5.51% Lateral Movement – TA0008
- 5.07% Resource Development – TA0042
- 4.88% Impact – TA0040
- 3.41% Command and Control – TA0011

Top 10 MITRE ATT&CK techniques detected by ThreatDown MDR in 2025



- 16.87% Network Service Discovery – T1046
- 16.14% PowerShell – T1059
- 12.01% Registry Run Keys/Startup Folder – T1547
- 9.55% Phishing – T1566
- 8.51% Masquerading – T1036
- 8.41% Malware – T1587
- 8.26% Disable or Modify Tools – T1562
- 7.88% User Execution – T1204
- 6.39% Disable or Modify System Firewall – T1562
- 5.98% De-obfuscate/Decode Files or Information – T1140

Top 10 MITRE ATT&CK tactics detected by ThreatDown MDR in 2025



1. Initial access

In attacks seen by ThreatDown in 2025, cybercriminals' favored methods of initial access were phishing (T1566), and software vulnerabilities in internet-facing security appliances—particularly Fortinet firewalls and SonicWall VPNs (T0819).

Five Fortinet and four SonicWall vulnerabilities were added to the CISA Known Exploited Vulnerabilities Catalog in 2025, and four of those were known to have been used in ransomware campaigns.

Phishing campaigns used familiar brands and believable lures like secure document downloads. Increasingly, attackers relied on AI-generated emails to eliminate the errors that many rely on to identify phishing and to produce more polished, convincingly personalized messages at scale. Using simple techniques such as checking MX records, attackers served victims fake versions of Google or OneDrive login screens tied to the victims' own domains. In some cases, victims were redirected to their real inboxes after harvesting credentials to minimize suspicion.

Protection layers that defend against initial access

Protection phase	Protection layer
Prevention	Email Security
	Vulnerability Assessment
	Patch Management
	DNS Filtering
	Firewall Management
Pre-delivery	Web Protection
	Browser Protection



2. Persistence

Many attackers now prefer to use remote monitoring and management (RMM) tools for persistence (MITRE T1219) rather than malware like remote access Trojans that might trigger detection from an endpoint protection platform (EPP).

RMM tools like Connect Wise and AnyDesk provide full keyboard access and do not look out of place on corporate networks. The portable version of AnyDesk can be run without a full installation, and some criminals have attempted to trick users into downloading and running it after successfully stealing their credentials in a phishing attack (MITRE T1078), in the same way they might once have tricked them into using malware.

Protection layers that defend against persistence

Protection phase	Protection layer
Prevention	Application Blocker
	Firewall Management
	DNS Filtering
Post-execution	Suspicious Activity Monitoring



3. Privilege escalation

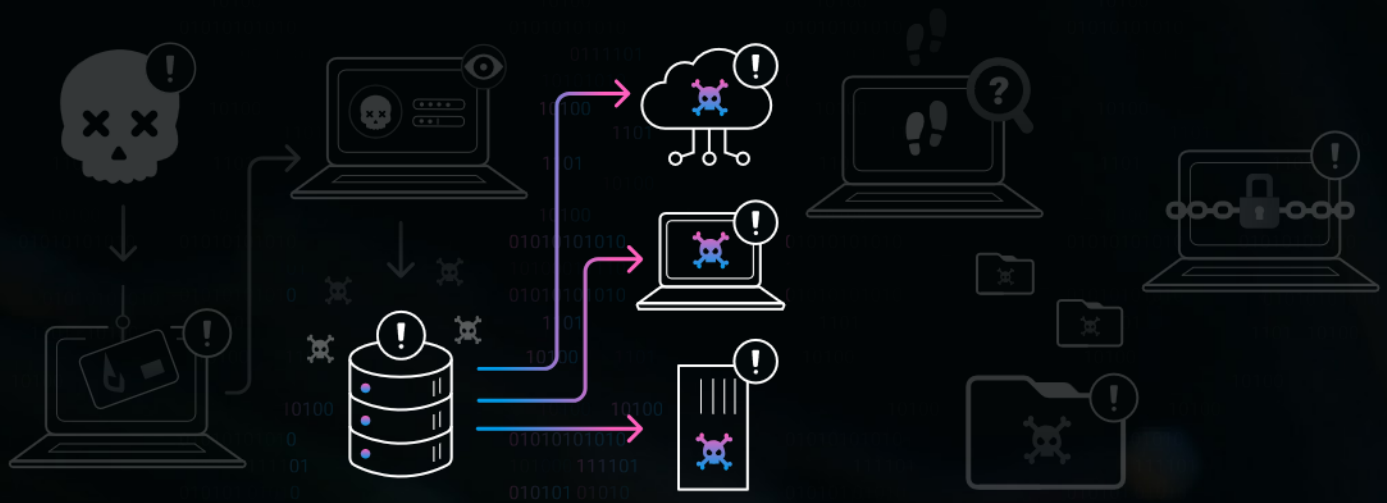
Attackers typically seek domain administrator privileges that give them full control over an organization's network. They achieve this through multiple pathways, often starting with exploitation of software vulnerabilities (MITRE T1068) to escalate from a standard user account to SYSTEM-level access by abusing vulnerable services, kernel components, or weak privilege separation.

After gaining elevated privileges, attackers commonly abuse Windows token mechanisms to deepen access. Token impersonation privileges can allow them to act as SYSTEM without stealing credentials (MITRE T1134). They may also hijack privileged service pipes, manipulate Kerberos authentication by stealing or forging tickets (MITRE T1558.001), or extract service account credentials through Kerberoasting (MITRE T1558.003).

Credential-stealing tools accelerate escalation and lateral reach. Mimikatz (MITRE S0002) extracts credentials from memory (MITRE T1003), while LaZagne (MITRE S0349) retrieves passwords stored locally on compromised systems (MITRE T1555).

Protection layers that defend against privilege escalation

Protection phase	Protection layer
Prevention	Vulnerability Assessment
	Patch Management
Pre-delivery	Application Hardening
	Exploit Mitigation
Pre-execution	Application Blocker
Post-execution	Suspicious Activity Monitoring
	Endpoint Isolation



4. Lateral movement

During lateral movement, attackers expand from a compromised machine into surrounding networks using native Windows tools and legitimate administrative utilities. They may execute commands via PowerShell (MITRE T1059.001) or Windows Management Instrumentation (MITRE T1047), pivot using built-in remote access such as PsExec (MITRE T1021.002) or Remote Desktop Protocol (MITRE T1021.001), or deploy post-exploitation frameworks like Cobalt Strike or Metasploit to automate movement. This "living off the land" approach is now standard in attacks against businesses.

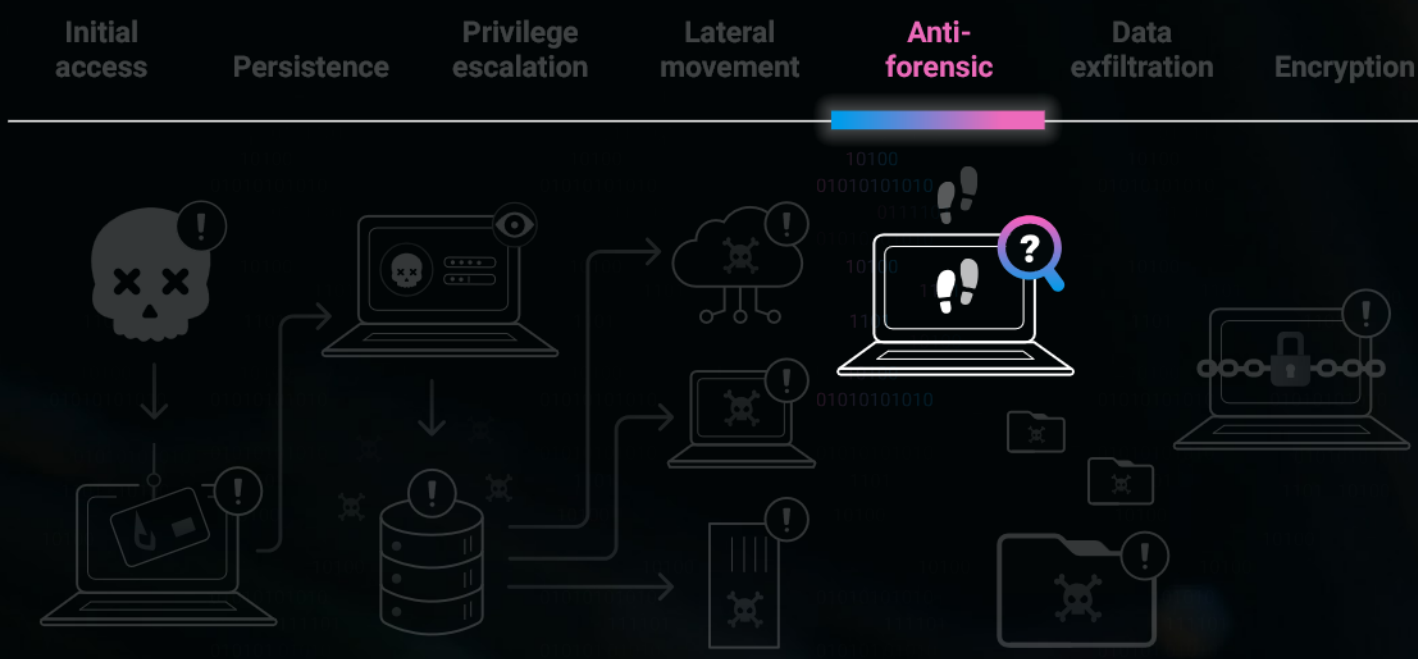
Attackers intensify credential harvesting during this phase to access additional systems, maintain persistence if privileged accounts are changed, and move through lower-privilege or service accounts to reduce noise and avoid detection. They also enumerate domain trusts (MITRE T1482) to identify connected domains or forests they can pivot into.

In addition, attackers may intercept and relay NTLM authentication attempts (MITRE T1557.002), enabling access to other machines without possessing valid credentials. Stolen credentials can also be used to pivot into cloud identities, federated SSO environments, service accounts, or API keys, extending the attack far beyond the initial host.

A common on-premise target is NTDS.dit (MITRE T1003.003), the Active Directory database. Attackers may extract it by abusing replication APIs (MITRE T1003.006), creating volume shadow copies or snapshots, or exporting hashes for offline cracking.

Protection layers that defend against lateral movement

Protection phase	Protection layer
Prevention	Firewall Management
Pre-delivery	Protocol Hardening
Post-execution	Suspicious Activity Monitoring
	Endpoint Isolation



5. Anti-forensics

Attackers often remove evidence of their activity before exfiltration or encryption. This may include clearing Windows event logs (MITRE T1070.001), deleting PowerShell histories, rotating or corrupting logs, or disabling audit policies.

By covering their tracks, attackers make incident response and root-cause analysis significantly more difficult, delaying containment efforts and reducing defenders' ability to understand what the attackers accessed or stole.

Protection layers that defend against anti-forensics

Protection phase	Protection layer
Post-execution	Suspicious Activity Monitoring
	Flight Recorder
	Endpoint Isolation



6. Data exfiltration

In attacks against businesses, the target is data, which is typically either stolen and held to ransom, encrypted and held to ransom, or both. During and after lateral movement, expect attackers to identify important data and exfiltrate it.

Attackers using an RMM tool may use its built-in file transfer or scripting features to collect and extract data under the guise of normal administrative activity (MITRE T1041). They may also transfer data to cloud platforms likely to be trusted by the target's firewall, such as Google Drive, OneDrive, or Dropbox (MITRE T1567.002).

Protection layers that defend against data exfiltration

Protection phase	Protection layer
Prevention	DNS Filtering
	Firewall Management
Pre-delivery	Web Protection
Post-execution	Suspicious Activity Monitoring
	Endpoint Isolation



7. Encryption

The natural end state of cyberattacks against businesses is ransomware with attackers encrypting data across the organization and demanding payment for a decryptor (MITRE T1486).

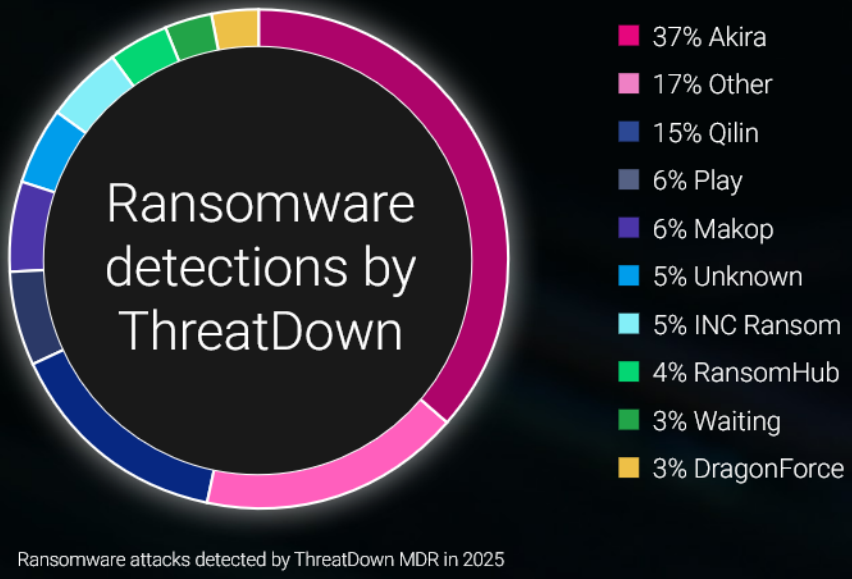
Before launching encryption, attackers routinely remove recovery mechanisms that could weaken their leverage. This commonly includes deleting Windows Volume Shadow Copies, purging cloud snapshots, or disabling backups using built-in tools and compromised administrative credentials (MITRE T1490).

In 2025, ThreatDown analysts observed attackers increasingly seeking out unprotected machines inside victim networks. These systems—often shadow IT devices, older servers, or endpoints lacking endpoint detection and response (EDR)—provide safe spaces to stage tooling, harvest credentials, and ultimately run encryption remotely.

> In 2025, 86% of encryption attacks were conducted remotely.

In 2025, ThreatDown analysts observed attackers increasingly staging encryption from unprotected systems inside target networks. These unmanaged or EDR-less machines provide safe locations to prepare tooling and execute encryption remotely. Attackers may mount drives from protected systems or issue remote commands using tools such as PsExec (MITRE T1021.002), causing the NT kernel on the target machine to perform the file operations.

Because the malicious process runs elsewhere, EDR may detect the effects of encryption but cannot quarantine the source. Until the originating system is identified and isolated, attackers can retry, escalate, or pivot to additional targets.



Ransomware attacks detected by ThreatDown MDR in 2025

Protection layers that defend against encryption

Protection phase	Protection layer
Post-execution	Ransomware Detection
	Suspicious Activity Monitoring
	Endpoint Isolation
	Ransomware Rollback

> "Almost all the ransomware cases we see come from IP addresses on the network that aren't protected."
>
> **– ThreatDown Threat Research Analyst**

Five key operating patterns

Modern cyberattacks prioritize stealth. As EDR coverage expanded, criminals developed techniques designed to suppress alerts or complete operations before defenders could respond. ThreatDown now identifies five stealthy operational patterns that small and mid-sized businesses should understand.

1. Faster attacks

There are strong incentives for cybercriminals to accelerate their operations, and dwell times have plummeted in recent years. The longer an attacker remains inside a network, the more time IT or security teams have to correlate and escalate the suspicious activity alerts generated by EDR.

By compressing multi-stage operations into hours, attackers reduce the number of alerts they generate and drastically shrink the window in which defenders can intervene. Speed is no longer a byproduct of automation; it is a deliberate strategy to stay ahead of human response.

2. Working at night

Attackers deliberately time their activity for periods when organizations are the least able to respond.

By operating during quiet periods such as nights, weekends, and holidays, attackers exploit natural gaps in organizational attention, when EDR alerts are more likely to go unnoticed or response times are at their slowest. This enables adversaries to move through critical stages of an intrusion before anyone intervenes, turning low staffing levels into a strategic advantage.

3. Living off the land

Attackers routinely use living off the land techniques to disguise their malicious activity. By relying on the tools, commands, and accounts that IT staff use every day, attackers make much of their behavior resemble routine administration rather than an obvious intrusion.

This familiarity makes it far harder for IT and security teams to distinguish harmful actions from legitimate work. The ambiguity gives attackers room to explore networks, escalate privileges, and stage their operations without attracting immediate attention. Living off the land shifts the defender's task from identifying malicious software to identifying malicious intent.

4. Staging from blind spots

Cybercriminals seek out blind spots in company networks that allow them to operate without surveillance. Unknown "shadow IT" systems, unmonitored endpoints, unsupported operating systems, and machines with overly broad EDR exclusions give attackers places where their activity is unlikely to be logged, analyzed, or blocked.

These blind spots let attackers harvest credentials, stage tools, map out networks, and launch remote commands without being detected. The less visibility defenders have, the more freedom attackers have.

Five key operating patterns

5. Attacking security and recovery software

Cybercriminals increasingly prioritize disabling the tools that could stop them or allow a business to recover. As defensive technologies improved, attackers responded by targeting the systems that provide visibility, control, and resilience. Removing these systems makes a successful attack more likely and eliminates alternatives to paying ransoms.

- **ESXi hypervisors**
  Attackers target ESXi hypervisors because compromising this layer gives them control over every virtual machine beneath it—along with the security agents and backup tools those machines depend on. Since ESXi hosts cannot run EDR, they represent a natural blind spot. By gaining control of the hypervisor, attackers remove a major source of visibility and gain the ability to disrupt entire environments from a single point.

- **EDR killers**
  EDR killers are a broad category of tools designed to blind or disable the EDR protection. But defeating modern EDR platforms is not trivial. One of the most widely used approaches is Bring Your Own Vulnerable Driver (BYOVD) where attackers load a legitimately signed but vulnerable driver to gain kernel-level privileges—the highest level of access an operating system will allow. A successful assault on EDR software gives attackers the freedom to operate without being exposed.

> To speed up attacks, Akira ransomware only partially encrypts files. Although not as thorough as fully encrypting every file, this approach causes significant damage while using a less conspicuous amount of CPU and enabling much quicker operation.

- **Deleting shadow copies and backups**
  Backups undermine ransom demands by giving organizations a way to clean up and restore compromised and encrypted systems. By wiping shadow copies, corrupting backup repositories, or deleting cloud snapshots, attackers leave businesses with fewer options, longer downtime, and greater pressure to pay.

Ransomware

8%
increase
in known ransomware attacks in 2025

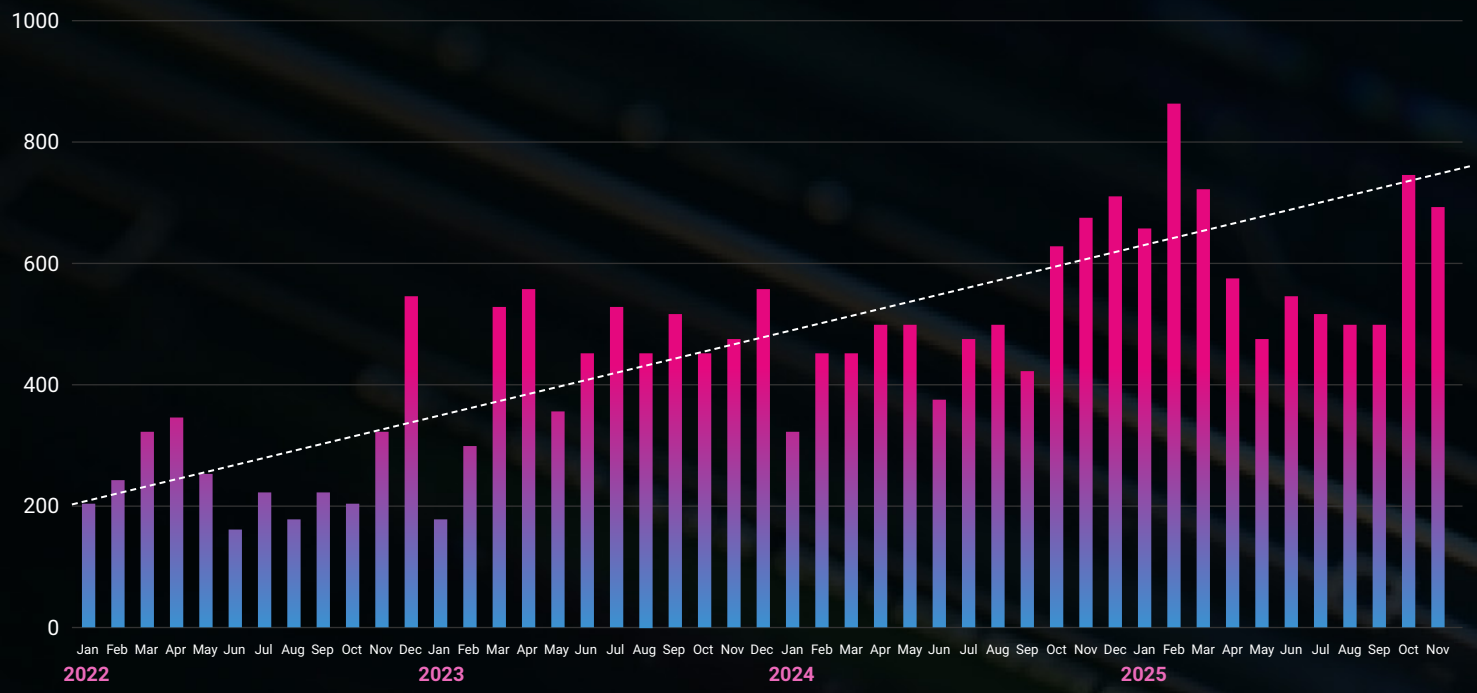
135
countries
experienced ransomware attacks in 2025

$2.5
billion
estimated cost of the Jaguar Land Rover attack

The clearest and most complete view we have of the criminal ecosystem targeting businesses comes from ransomware, where attackers use their dark web leak sites to broadcast information about their attacks.

2025 was the worst year for ransomware on record. Attacks increased 8% year over year, driven by the two worst months ever recorded: February and October.

Known ransomware attacks by month



Ransomware

Geography

The US remained the most attacked country in the world by a large margin, accounting for almost half of all known ransomware events in 2025. Elsewhere, attacks were heavily concentrated in other English-speaking economies and Western Europe.



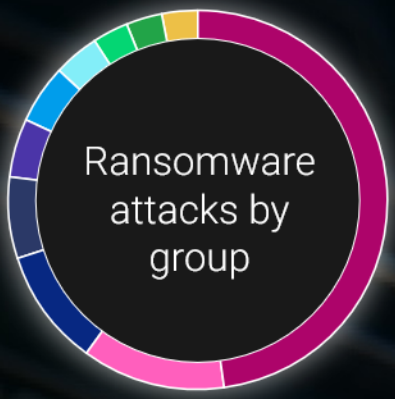
- 48% United States
- 27% Other
- 5% Canada
- 5% Germany
- 4% United Kingdom
- 3% France
- 2% Italy
- 2% Spain
- 2% India
- 2% Brazil

Companies from Russia, China, and much of the Global South were largely absent from leak sites. This pattern reflects long-standing geopolitical and economic dynamics in the ransomware ecosystem: Cybercriminals focus on wealthier economies with familiar technology stacks and languages, and where political or law-enforcement blowback is minimal.

Ransomware groups

Over the last few years, the growth of ransomware has been driven by a steady increase in the number of active groups. Most gangs are small, conducting only one or two attacks per month, and the collective impact of these smaller operators has grown over time.

While a handful of prolific groups still account for a disproportionate share of victims, the era of a single brand like LockBit or Conti dominating the ransomware ecosystem is over. Repeated law-enforcement takedowns, arrests, infrastructure seizures, and internal disputes have fractured large operations and driven affiliates to migrate between programs. At the same time, the continued expansion of ransomware-as-a-service models has lowered barriers to entry, enabling smaller and increasingly region-specific groups to emerge, rebrand, and operate with greater resilience. The result is a ransomware landscape that is structurally stable but increasingly unpredictable and spread across more threat actors.



- 48% Other
- 12% Qilin
- 10% Akira
- 7% Cl0p
- 5% Play
- 5% INC Ransom
- 4% SafePay
- 3% Lynx
- 3% RansomHub
- 3% DragonForce

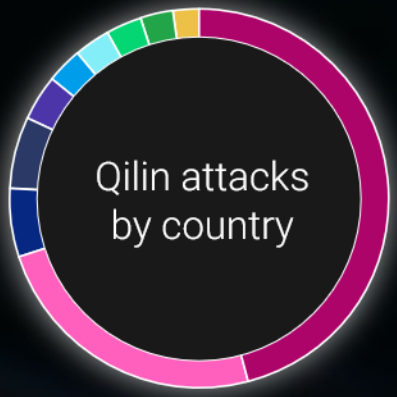
Ransomware

Qilin

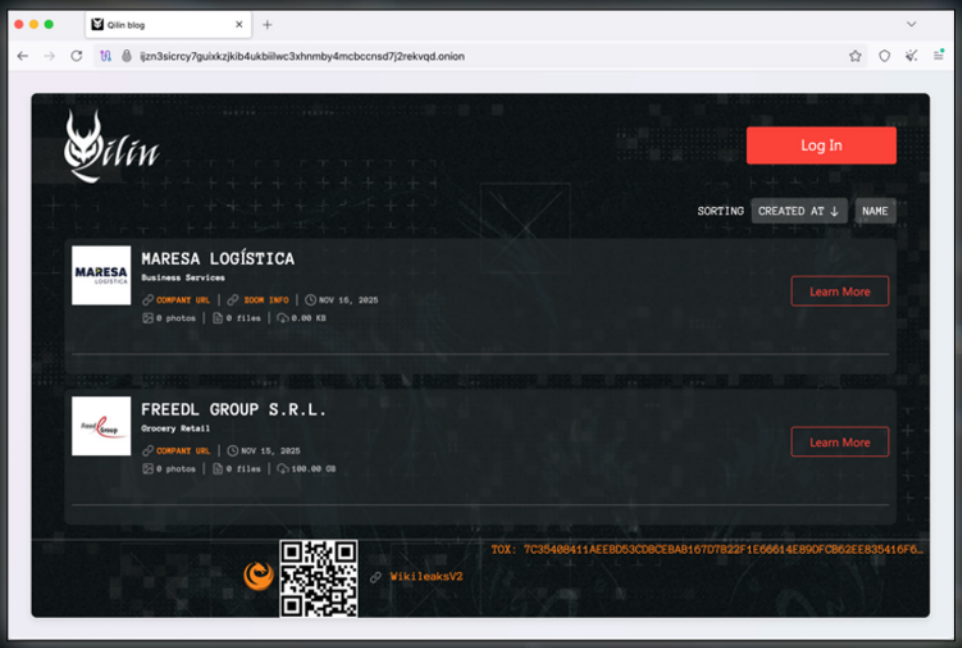
Qilin surged to become the most active ransomware group of 2025, carrying out more attacks than any other two groups combined. Its rise suggests a mature affiliate model, strengthened by the collapse or fragmentation of LockBit, ALPHV, and RansomHub.

Qilin targets mid-sized companies, regional firms, and local service providers across manufacturing, business services, healthcare, construction, and technology rather than large multinationals.

Targets:	Mid-sized organizations
Extortion model:	Encryption, data theft
Key tactics:	Exploitation of unpatched VPNs and remote access appliances, credential harvesting



- 46% United States
- 24% Other
- 6% Canada
- 6% France
- 4% Spain
- 3% United Kingdom
- 3% Japan
- 3% Germany
- 3% Italy
- 2% Republic of Korea



Ransomware

Akira

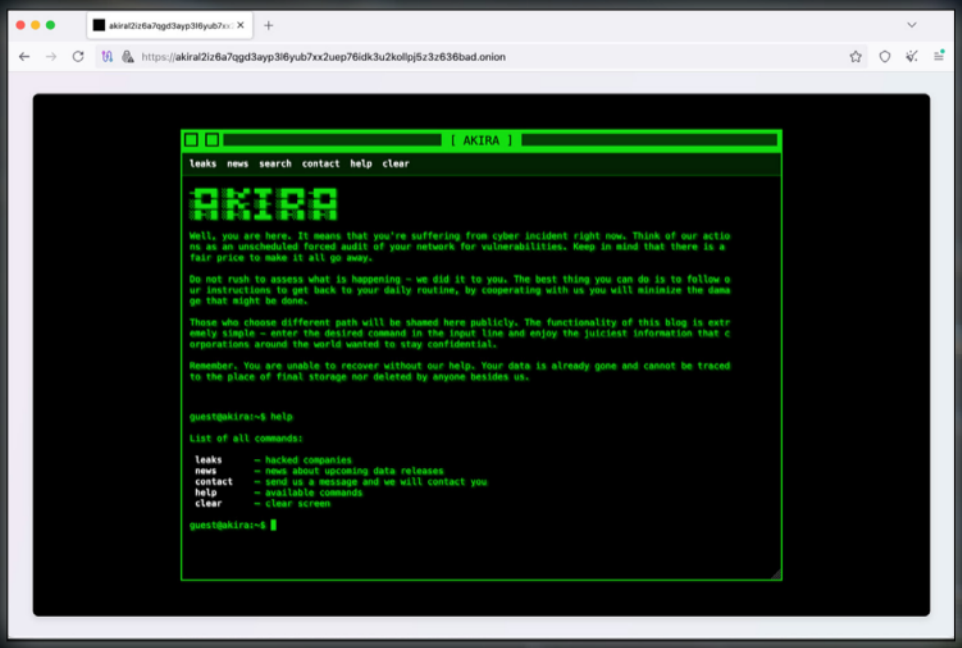
Akira was the second most active ransomware group in 2025 with activity concentrated in the United States, Germany, Canada, and Italy.

Akira overwhelmingly targets small and mid-sized organizations, including regional service providers, mid-market manufacturers, local public sector entities, and modest technology firms.



Targets:	SMEs and mid-sized organizations
Extortion model:	Encryption, data theft
Key tactics:	Exploitation of unpatched VPNs, stolen credentials

- 61% United States
- 15% Other
- 5% Germany
- 5% Canada
- 4% Italy
- 3% Spain
- 2% United Kingdom
- 2% Switzerland
- 2% Brazil
- 1% Australia



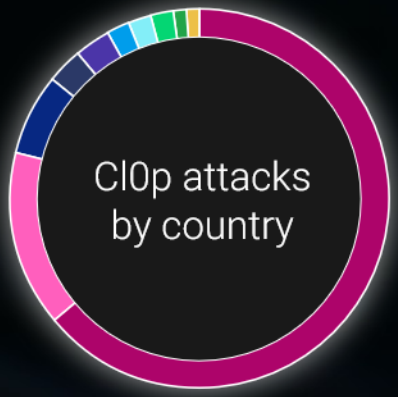
Ransomware

Cl0p

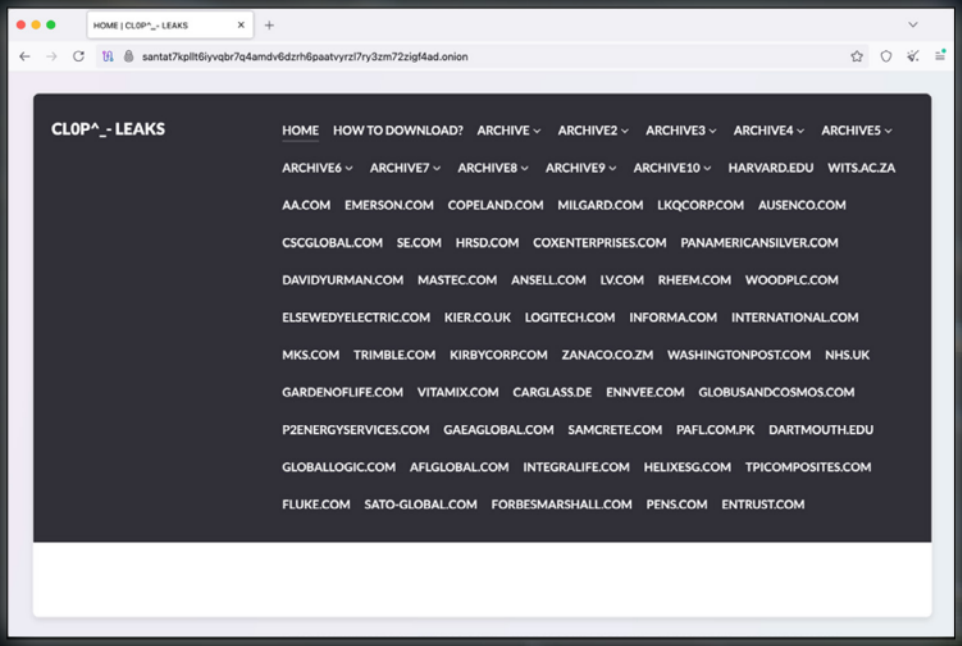
Cl0p remained one of the most strategically significant groups in 2025, conducting sporadic, high-impact mass-exploitation campaigns using zero-day vulnerabilities.

While it does hit mid-market organizations, Cl0p's most consequential operations disproportionately affect larger enterprises and complex environments running the file-transfer technologies it targets.

Targets:	Enterprises
Extortion model:	Data theft
Key tactics:	Zero-day exploitation of file-transfer and data-movement systems



- 64% United States
- 15% Other
- 7% Canada
- 3% Germany
- 3% United Kingdom
- 2% Mexico
- 2% India
- 2% Japan
- 1% Australia
- 1% France



Ransomware

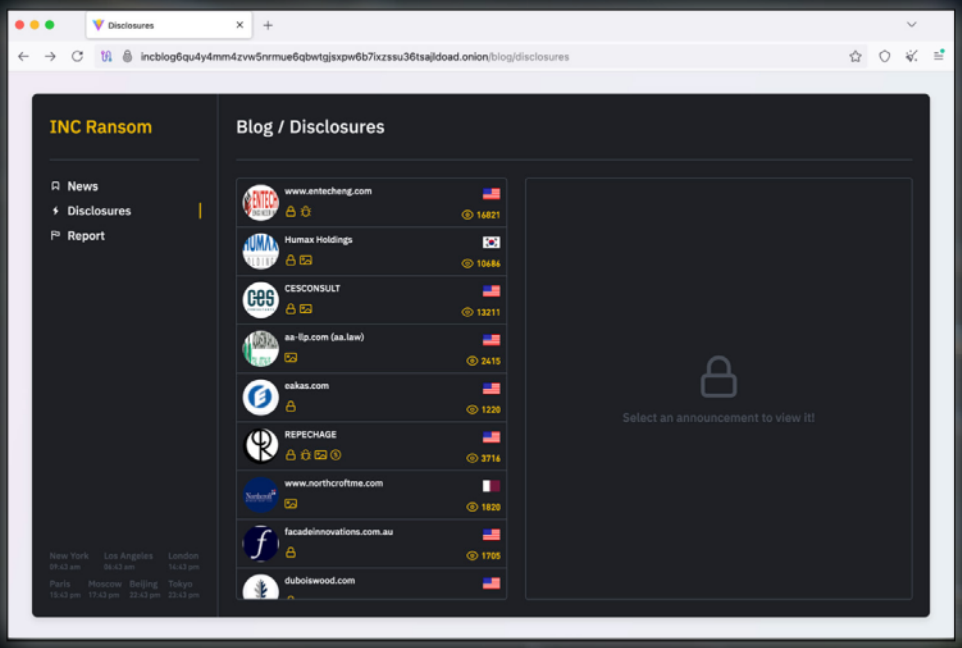
INC Ransom

INC Ransom was the fastest-rising ransomware group of 2025, nearly doubling its activity to become the fourth most active group . This growth reflects an expanding affiliate model, which enabled the group to scale operations across multiple campaigns. That expansion was most visible in attacks against small and mid-sized organizations.



Targets:	SMEs and mid-sized organizations
Extortion model:	Encryption, data theft
Key tactics:	Compromised credentials

- 54% United States
- 19% Other
- 8% Canada
- 6% Germany
- 4% United Kingdom
- 2% Australia
- 2% Austria
- 2% Italy
- 2% France
- 1% Brazil



How AI is reshaping cybercrime

2025 marked the moment when cybercrime began its shift toward an AI-driven future. Hands-on-keyboard intrusions still dominated, but the year delivered the first confirmed cases of AI-orchestrated attacks—alongside deepfake-enabled social engineering and AI agents that outperformed humans at discovering vulnerabilities.

The constraints that limit human attackers are beginning to fall away as operations are delegated to near-infinitely scalable AI agents that learn and adapt.

The following sections examine four developments from 2025 that reveal how AI will reshape cybercrime in 2026.

Deepfakes

2025 saw an escalating series of warnings from regulators, financial institutions, law-enforcement agencies, and industry bodies highlighting deepfakes as an active and rapidly growing threat to businesses.[10] In July, OpenAI CEO Sam Altman added urgency to these concerns, stating that generative AI had already "fully defeated" the face and voice authentication systems used by banks and was creating "a significant impending fraud crisis."[11]

A 2025 analysis by IBM found that AI played a role in 16% of breaches with deepfake voice or video manipulation accounting for 35% of those incidents.[12] Several high-profile fraud cases—including attacks against WPP,[13] an unnamed cryptocurrency firm,[14] and the $25 million social-engineering theft from global engineering firm Arup[15]—demonstrated that criminals can now create impersonations convincing enough to succeed in live video calls.

> AI played a role in 16% of breaches with deepfake voice or video manipulation accounting for 35% of those incidents.

Deepfake voice, image, and video impersonation now requires minimal expertise and only a handful of reference images or seconds of audio. Criminals are using these capabilities across a wide spectrum of attacks: creating fabricated IDs for financial fraud; mimicking IT or helpdesk staff to persuade employees to share passwords, reset multi-factor authentication (MFA), or approve remote access; and impersonating executives to conduct highly convincing forms of CEO fraud.

ThreatDown expects AI-driven social engineering operations to scale significantly throughout 2026 and to emerge as the dominant form of social engineering used by attackers.

[10] Federal Reserve (2025), Deepfakes and the AI Arms Race in Bank Cybersecurity, https://www.federalreserve.gov/newsevents/speech/barr20250417a.htm
[11] C-SPAN, OpenAI CEO Sam Altman Speaks at Federal Reserve Conference, https://www.c-span.org/program/public-affairs-event/openai-ceo-sam-altman-speaks-at-federal-reserve-conference/662859
[12] IBM (2025), Cost of a Data Breach Report 2025, https://www.ibm.com/reports/data-breach
[13] Financial Times (2024), WPP boss targeted by deepfake scammers using voice clone, https://www.ft.com/content/308c42af-2bf8-47e4-a360-517d5391b0b0
[14] Huntress (2025), Feeling Blue(Noroff): Inside a Sophisticated DPRK Web3 Intrusion, https://www.huntress.com/blog/inside-bluenoroff-web3-intrusion-analysis
[15] The Guardian, UK engineering firm Arup falls victim to £20m deepfake scam, https://www.theguardian.com/technology/article/2024/may/17/uk-engineering-arup-deepfake-scam-hong-kong-ai-video

How AI is reshaping cybercrime

Vulnerability discovery

In 2025, vulnerability research reached a turning point as AI systems like Google's Big Sleep, XBOW, and Auto Exploit—a publicly demonstrated AI-driven exploit pipeline—began to outperform human experts. By mid-2025:

- **Big Sleep** reported 20 previously unknown flaws in open-source projects—including a memory-safety bug in SQLite that had previously eluded detection.[16]
- **The autonomous XBOW agent** eclipsed its human competitors on HackerOne to be crowned the most prolific bug bounty hunter in the USA.[17]
- **Auto Exploit** was generating proof-of-concept exploit code in as little as 15 minutes, compressing timelines that previously took hours or days.[18]

All three systems are working for the benefit of defenders, but their existence carries an implicit warning: AI agents are now capable of surfacing exploitable bugs at a speed and scale that far outstrips human analysts.

This shift has profound implications. As AI systems become more widely available, the number of actors capable of identifying exploitable vulnerabilities will expand dramatically. Zero-day discovery, once bottlenecked by the scarcity of expert talent, will become a matter of computer power and model quality.

ThreatDown expects cybercriminals to embrace AI-discovered exploits in 2026, increasing the likelihood of continuous, automated scrutiny of internet-facing appliances and significantly shorter patch windows. As Auto Exploit's creator Efi Weiss warned, "Defenders will need to be drastically faster…"

> "For the first time in bug bounty history, an autonomous penetration tester has reached the top spot on the US leaderboard."
>
> **– Nico Waisman, CISO, XBOW**

[16] TechCrunch (2025), Google says its AI-based bug hunter found 20 security vulnerabilities, https://techcrunch.com/2025/08/04/google-says-its-ai-based-bug-hunter-found-20-security-vulnerabilities/
[17] XBOW (2025), The road to Top 1: How XBOW did it, https://xbow.com/blog/top-1-how-xbow-did-it
[18] Efi Weiss (2025), Can AI weaponize new CVEs in under 15 minutes? https://valmarelox.substack.com/p/can-ai-weaponize-new-cves-in-under

How AI is reshaping cybercrime

Autonomous ransomware

One of the most significant cybersecurity developments of 2025 was the first in-the-wild use of autonomous AI agents to scale extortion operations. Historically, the biggest global ransomware groups struggled to scale beyond a few hundred attacks per month, limited by the need for skilled operators performing hands-on-keyboard intrusions.

In the 2025 State of Malware report, ThreatDown predicted that ransomware groups would adopt AI agents to break this scalability barrier. Eight months later, the first signs of that shift appeared in Anthropic's August 2025 Threat Intelligence report, which detailed an extortion campaign (GTG-2002) where a single criminal used Claude Code to automate and scale attacks against targets in healthcare, defense, emergency services, and local government.

The agent conducted reconnaissance across thousands of VPN endpoints, harvested credentials, penetrated internal networks, and analyzed stolen data. Claude Code also acted as an advisor, creating per-victim extortion strategies and generating tailored ransom notes based on real-time assessments of each organization's financial and regulatory exposure.[19]

> Claude Code also acted as an advisor, creating per-victim extortion strategies and generating tailored ransom notes.

Anthropic concluded that GTG-2002 demonstrated that "a single operator can achieve the impact of an entire cybercriminal team through AI assistance." The company also warned that AI-driven intrusions will become harder to defend against as "AI-generated attacks adapt to defensive measures in real-time."

GTG-2002 was the first evidence outside a research lab that cybercriminals are beginning to offload core intrusion tasks to AI agents. ThreatDown expects that in 2026, these emerging capabilities will mature into fully autonomous ransomware pipelines that allow individual operators and small crews to attack multiple targets simultaneously at a scale that exceeds anything seen in the ransomware ecosystem to date.

[19] Anthropic (2025), Detecting and countering misuse of AI: August Detecting and countering misuse of AI: August 2025, https://www.anthropic.com/news/detecting-countering-misuse-aug-2025.

How AI is reshaping cybercrime

Malicious MCP servers

The most concerning development of 2025 for businesses was the marriage of highly autonomous AI with powerful penetration testing tools, via the Model Context Protocol (MCP)—an open standard that lets AI models connect to external tools, data sources, and applications.

2025 saw a rapid progression in AI-driven red teaming as researchers used MCP to bind AI agents to individual tools like nmap and then to fully featured red-team frameworks such as Mythic. This culminated in MIT research released in November showing that MCP could underpin stealthy, autonomous, multi-agent attack swarms capable of completing end-to-end compromises with little to no human interaction, in under an hour.[20]

In the same month, it became clear that criminals had been researching similar techniques. An Anthropic threat intelligence report provided the first real-world example of this architecture being used in active attacks. Investigators uncovered a cybercrime operation (GTG-1002) that paired the Claude Code agent with MCP-enabled tooling to conduct vulnerability discovery, exploit development, access validation, credential harvesting, persistence, lateral movement, and structured intelligence reporting near-autonomously across 30 organizations.[21] MCP servers coordinated the underlying tools and maintained operational state, while human operators intervened only to approve major steps or redirect priorities.

Although some security researchers expressed skepticism about Anthropic's claims, the MIT paper demonstrated that the decoupled MCP-based C2 architecture and multi-agent coordination attributed to GTG-1002 is technically feasible.[22]

The convergence of open-source offensive tooling, MCP wrappers, and highly capable AI agents puts powerful red-teaming capabilities in the hands of defenders, while also creating a path for cyberattacks that are faster, more adaptive, and far more scalable than anything achievable through hands-on-keyboard intrusions.

ThreatDown expects that in 2026, MCP-based attack frameworks will become a defining capability of cybercriminals targeting businesses. The shift from AI-assisted to AI-executed operations will reduce the need for skilled operators, lower barriers for less experienced attackers, and allow well-resourced groups to scale their campaigns far beyond current limits. Vulnerability scanning, exploitation, lateral movement, and data theft will increasingly be carried out by agents operating continuously and at machine speed, which will bring about a step-change in the tempo and efficiency of cybercrime.

[20] Smith et al. (2025), Hiding in the AI Traffic: Abusing MCP for LLM-Powered Agentic Red Teaming, arXiv preprint, https://arxiv.org/abs/2511.15998
[21] Anthropic (2025), Disrupting the first reported AI-orchestrated cyber espionage campaign, https://www.anthropic.com/news/disrupting-AI-espionage
[22] TechRadar (2025), Experts cast doubt over Anthropic claims that Claude was hijacked to automate cyberattacks, https://www.techradar.com/pro/security/experts-cast-doubt-over-anthropic-claims-that-claude-was-hijacked-to-automate-cyberattacks

Conclusion

2025 marked a decisive turning point for global cybersecurity. As ransomware escalated and hands-on-keyboard intrusions dominated attacks on businesses, the first wave of AI-orchestrated operations signaled an irreversible acceleration toward a post-human future. AI agents executed complex, multi-stage compromises with minimal supervision; researchers showed that AI-driven pipelines can go from patch to exploit in 15 minutes; and an AI agent outperformed every human bug bounty hunter to go top of HackerOne's global leaderboard.

As these capabilities mature in 2026, they will rapidly diffuse through the cybercrime ecosystem. Defenders will face adversaries capable of continuous, intelligent, adaptable activity across many targets at once, unconstrained by fatigue or staffing limitations. Malicious agents will use the same playbooks as human hackers but execute them at machine speed.

For businesses, the urgency is clear. Traditional reliance on slow patching cycles, limited visibility, and reactive security will no longer be sufficient. Organizations must shrink their attack surfaces, harden identity systems, close blind spots, accelerate remediation, and adopt continuous monitoring and expert-led services such as Managed Detection and Response to keep pace.

Cybercrime has entered its machine-scale era. The organizations that invest now in visibility, resilience, and automation will be best positioned to adapt to a threat landscape evolving faster than ever before.

How to protect your company

Whether they are conducted by a human or an artificial intelligence, modern intrusions rely on legitimate tools and credentials, progress through multiple paths, and use hands-on tactics to adapt to defensive countermeasures. ThreatDown's award-winning protection addresses today's evolving cybersecurity challenges with a stack of integrated, overlapping security layers that offer protection across the entire attack cycle: from attack surface reduction; to prevention, detection, and response; and full remediation.

ThreatDown protection layers provide security across the entire attack cycle

Protection phase	Protection layer
Prevention	Email Security
	Vulnerability Assessment
	Patch Management
	DNS Filtering
	Firewall Management
Pre-delivery	Web Protection
	Browser Protection
	Protocol Hardening
	Application Hardening
	Payload Analysis
	Cloud Sandbox
Pre-execution	Application Blocker
	Exploit Mitigation
Post-execution	Application Behavior
	Suspicious Activity Monitoring
	Anomaly Detection (Machine Learning)
	Ransomware Detection
	Endpoint Isolation
	Active Response Shell
	Ransomware Rollback
	Remediation Linking Engine
	Flight Recorder

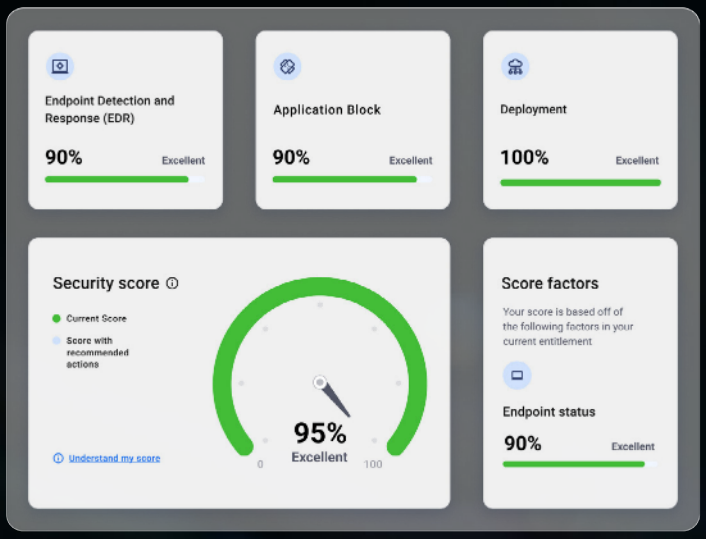
How to protect your company

ThreatDown Managed Detection & Response (MDR)

ThreatDown MDR, delivers around-the-clock threat detection, investigation, and remediation by elite security analysts who specialize in stopping ransomware and other advanced threats before damage is done.

Organizations face mounting pressure to stay secure with limited staff, expertise, and time. ThreatDown MDR is built to close that gap. You'll experience always-on coverage designed to deliver peace of mind and support your business continuity at every turn.

- **24x7x365 monitoring**
  ThreatDown's expert MDR team watches your endpoints day and night, including weekends and holidays, so you can rest easy.
- **Powered by award-winning EDR**
  The ThreatDown Endpoint Detection & Response (EDR) platform detects and blocks threats with advanced layers of protection, persistent ransomware rollback, and deep malware cleanup.
- **Backed by SIEM and SOAR**
  Detection data is enriched and prioritized in real time using integrated threat intelligence, automation, and correlation to reduce noise and accelerate response.
- **Flexible response options**
  Choose hands-on analyst-led remediation or actionable guidance for your team, whichever fits your operating model best.



- **Fast deployment, full visibility**
  Get up and running in minutes with lightweight agents and intuitive onboarding. ThreatDown's dashboards give you a clear view of what's happening, what has been stopped, and what actions were taken.
- **Affordable and scalable**
  ThreatDown's pricing is transparent and cost-effective. Whether you're supporting 50 endpoints or 5,000, ThreatDown MDR scales with your needs.

Proven, recognized protection
ThreatDown MDR is backed by industry validation:

THREATDOWN™

Copyright © 2026, ThreatDown. All rights reserved. ThreatDown and the ThreatDown logo are trademarks of ThreatDown. Other marks and brands may be claimed as the property of others. All descriptions and specifications herein are subject to change without notice and are provided without warranty of any kind. 01/26

